Application of Chunking to the Design of Complex Information Displays

by

Michael J. Kahn

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

in

Industrial Engineering and Operations Research

APPROVED:

R. J. Beaton, Co-Chairman

H. L. Snyder, Co-Chairman

G. J. Buhyoff

R. D. Dryden

P. T. Kemmerling

April, 1990

Blacksburg, Virginia

c.2

LD
5655
V856
1990
K336
c.2

Application of Information Chunking to the

Design of Complex Information Systems

(ABSTRACT)

by

Michael J. Kahn

Committee Co-Chairmen: Robert J. Beaton, Harry L. Snyder

Industrial Engineering and Operations Research

Chunking is defined as the process of organizing information into multi-item clusters appropriate to solving a particular task. It was hypothesized that dynamic information displays could be formatted to group related items. This formatting would facilitate chunking and, in turn, serve to reduce task difficulty. Two novel display grouping formats were evaluated. In Tight Spatial Proximity, iconic symbols belonging to the same class were rearranged so that they were located proximally on the display. In Chunks-In-Sequence (CIS), symbol classes were presented sequentially, but all members within a symbol class were presented simultaneously.

Ten individuals participated in Experiment 1. During a trial, 2 to 25 symbols (e.g., cannon, radar, etc.) were presented briefly on a computer display. Displayed symbols were sampled from a larger symbol set (i.e., 64 total symbols) representing eight unique symbol classes (e.g., military, highway signs, etc.). Participants then viewed a prompt and responded "yes" or "no" to indicate whether the probe symbol was a member of a class that had been presented during that trial. Results indicated that grouping formats helped participants chunk symbols into appropriate classes and solve the classification task with reduced error rates and subjective workload.

Experiment 2 was similar to Experiment 1, with the exception that participants were tasked to remember each symbol presented during a trial rather than just the symbol class names. This task required forming chunks, storing them in memory, and then parsing the chunks into component items. Results indicated that display grouping formats reduced subjective workload, but not error rates.

Considering both experiments, it is concluded that spatially or temporally grouped symbols can be chunked more easily than information displayed in a non-grouped format. However, grouping formats will not help operators parse these information chunks given detailed component recall demands.

ACKNOWLEDGEMENTS

This research was funded by contract F33615-88-D-0532 from the U. S. Air Force Armstrong Aerospace Medical Research Laboratory.

I would first like to thank Dr. Robert Beaton, who spent many hours working closely with me. His efforts contributed greatly to the quality of this dissertation and helped me develop as a human factors engineer. I also thank Dr. Harry Snyder for his professional and emotional guidance in this project and across my years at Virginia Tech. I thank committee members Dr. Gregory Buhyoff, Dr. Robert Dryden, and Professor Paul Kemmerling for their commitment to the process of education and for their compassion.

Finally, I thank my parents, Barbara and Sherwood Kahn, for imparting in me a love of knowledge.

TABLE OF CONTENTS

INTRODUCTION 1

BACKGROUND 3
STM Scanning Rate 3
STM and Cognitive Workload 4

CHUNKING 5
Conceptual Versus Positional Chunking 10
Chunking and Response Latency 13
Encoding times 13
Recall latency 14
Cavanagh (1972). 14
Presentation Order 14
Application of Chunking Principles to Visual Display 17
Display Design Incorporating Chunking 20
Efficiency and appropriateness. 20

RESEARCH OBJECTIVES 22

EXPERIMENT 1: METHODS 24
Participants 24
Apparatus, Stimuli, and Experimental Environment 24
Task 25
Design 27
Number Of Classes. 27
Number Of Items 27
Spatial Proximity 27
Presentation Order 29
Response Type 29
Blocks. 36
Dependent Measures 36
Procedure 36

EXPERIMENT 1: RESULTS AND DISCUSSION 37
Data Analysis Overview 38
SWAT Ratings 38
Effects due to display complexity 39
Effects due to display grouping 44
Combined effects of complexity and grouping 49
Effects due to Type 54
Error Rate Data 58
Effects due to display complexity 58
Effects due to display grouping 63
Combined effects of complexity and grouping 63
Effects due to Type 69

Response Time 71
Comments on the response time metric 71
Effects due to display complexity 74
Effects due to display grouping 74
Combined effects of complexity and grouping 74
Effects due to Type 78

EXPERIMENT 2: METHODS 82
Participants 82
Apparatus, Stimuli, and Experimental Environment 82
Task 82
Design 83
Number Of Classes. 83
Number Of Items 83
Spatial Proximity 83
Presentation Order 83
Response Type 83
Blocks. 84
Dependent Measures 84
Procedure 84

EXPERIMENT 2: RESULTS AND DISCUSSION 85
Data Analysis Overview 86
SWAT Ratings 86
Effects due to display complexity. 87
Effects due to display grouping 91
Combined effects of complexity and grouping 94
Effects due to Type 101
Error Rate 104
Effects due to display complexity 104
Effects due to display grouping 104
Combined effects of complexity and grouping 111
Effects due to Type 111
Response Time 112
Effects due to display complexity 112
Effects due to display grouping 112
Combined effects of complexity and grouping 112
Effects due to Type 119

CONCLUSIONS 121
Summary 121
Research Findings 122
Design Implications 123
Chunking versus parsing chunks 123
Chunkable versus non-chunkable information 123
CIS versus positional chunking 124
Positive versus negative searches. 124
Research Issues 125

REFERENCES 127

APPENDICES

Appendix A. SYMBOL BIT MAPS 130

Appendix B. PARTICIPANT INFORMED CONSENT FORM..... 139

Appendix C. BRIEF EXPLANATION OF SWAT..... 143

Appendix D. SWAT CARD SORT INSTRUCTIONS TO PARTICIPANTS..... 146

Appendix E. PARTICIPANTS' INSTRUCTIONS FOR EXPERIMENT 1..... 150

Appendix F. TABLES OF RESULTS FOR EXPERIMENT 1 152

Appendix G. PARTICIPANTS' INSTRUCTIONS FOR EXPERIMENT 2..... 178

Appendix H. TABLES OF RESULTS FOR EXPERIMENT 2 180

VITA..... 204

LIST OF FIGURES

Literature Review

Figure 1. A conceptual model of chunking 6

Figure 2. Learning curves of master, class A player, and beginner 8

Figure 3. Perceptual factors postulated in Gestalt psychology 11

Figure 4. Parallel versus Serial Presentation 16

Experiment Methods

Figure 5. Graphical Representation of Independent Variables for Experiments 1 and 2) 28

Figure 6. Display Showing Parallel-Presentation-Order/Random-Proximity 30

Figure 7. Display Showing Parallel-Presentation-Order/Medium-Proximity 31

Figure 8. Display Showing Parallel-Presentation-Order/Tight-Proximity 32

Figure 9. Display Showing CIS-Presentation-Order/Random-Proximity, first 0.8 s 33

Figure 10. Display Showing CIS-Presentation-Order/Random-Proximity, second 0.8 s 34

Figure 11. Display Showing CIS-Presentation-Order/Random-Proximity, third 0.8 s 35

Experiment 1: Graphs of SWAT Results

Figure 12. Main effect of Number Of Classes 40

Figure 13. Main effect of Number Of Items 41

Figure 14. Number Of Classes X Number Of Items 42

Figure 15. Number of symbols 43

Figure 16. Presentation Order 46

Figure 17. Proximity 47

Figure 18. Presentation Order X Proximity 48
Figure 19. Number Of Items X Presentation Order 50
Figure 20. Number Of Items X Proximity..... 52
Figure 21. Number Of Items X Presentation Order X Proximity 53
Figure 22. Number Of Classes X Number Of Items X Proximity 55
Figure 23. Type..... 56
Figure 24. Number Of Items X Type 57
Figure 25. Number Of Classes X Presentation Order X Type 59

Experiment 1: Graphs of Error Rate Results

Figure 26. Number Of Classes..... 60
Figure 27. Number Of Items..... 61
Figure 28. Number Of Classes X Number Of Items 62
Figure 29. Number of symbols 64
Figure 30. Proximity 65
Figure 31. Presentation Order X Proximity 66
Figure 32. Number Of Classes X Presentation Order..... 67
Figure 33. Number Of Items X Presentation Order 68
Figure 34. Number Of Classes X Presentation Order X Proximity 70
Figure 35. Number Of Items X Type. 72
Figure 36. Number Of Classes X Presentation Order X Type 73

Experiment 1: Graphs of Response Time Results

Figure 37. Number Of Classes..... 75
Figure 38. Number Of Classes X Proximity 76
Figure 39. Number Of Classes X Number Of Items X Presentation Order 77

Figure 40. Type 79
Figure 41. Number Of Classes X Type 80
Figure 42. Number Of Classes X Number Of Items X Type 81

Experiment 2: Graphs of SWAT Results

Figure 43. Number Of Classes 88
Figure 44. Number Of Items 89
Figure 45. Number Of Classes X Number Of Items 90
Figure 46. Number of symbols 92
Figure 47. Presentation Order 93
Figure 48. Proximity 95
Figure 49. Presentation Order X Proximity 96
Figure 50. Number Of Classes X Presentation Order 97
Figure 51. Number Of Items X Presentation Order 99
Figure 52. Number Of Items X Proximity 100
Figure 53. Number Of Items X Presentation Order X Proximity 102
Figure 54. Type 103

Experiment 2: Graphs of Error Rate Results

Figure 55. Number Of Classes 105
Figure 56. Number Of Items 106
Figure 57. Number Of Classes X Number Of Items 107
Figure 58. Number of symbols 108

Experiment 2: Graphs of Response Time Results

Figure 59. Number Of Classes 113

Figure 60. Number Of Items 114
Figure 61. Number of symbols 115
Figure 62. Number Of Classes X Presentation Order 116
Figure 63. Number Of Items X Presentation Order. 117
Figure 64. Number Of Classes X Number Of Items X Presentation Order 118
Figure 65. Type 120

LIST OF TABLES

Table 1. Cognitive Subtasks Completed when using military maps 19

Table 2. Symbol Set used in Experiments 1 and 2 26

Experiment 1: Tables of Results (in Appendix F)

Table F1. ANOVA Summary Table for SWAT Rating 153

Table F2. ANOVA Summary Table for Error Rate 155

Table F3. ANOVA Summary Table for Response Time 157

Table F4. Newman-Keuls Test of Number Of Classes, SWAT Rating 159

Table F5. Newman-Keuls Test of Number Of Items, SWAT Rating 160

Table F6. Newman-Keuls Test of Number Of Classes X Number Of Items, SWAT Rating 161

Table F7. Newman-Keuls Test of Proximity, SWAT Rating 162

Table F8. Newman-Keuls Test of Presentation Order X Proximity, SWAT Rating 163

Table F9. Newman-Keuls Test of Number Of Items X Presentation Order, SWAT Rating 164

Table F10. Newman-Keuls Test of Number Of Items X Proximity, SWAT Rating.... 165

Table F11. Simple-Effect F-tests for Presentation Order X Proximity for Each Level of Number Of Items, SWAT Rating 166

Table F12. Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 1, SWAT Rating 167

Table F13. Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 3, SWAT Rating 168

Table F14. Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 5, SWAT Rating 169

Table F15. Newman-Keuls Test of Number Of Items X Type, SWAT Rating 170

Table F16. Newman-Keuls Test of The Number Of Classes, Error Rate 171

Table F17. Newman-Keuls Test of Number Of Classes X Number Of Items, Error Rate 172

Table F18. Newman-Keuls Test of Proximity, Error Rate 173

Table F19. Newman-Keuls Test of Presentation Order X Proximity, Error Rate 174

Table F20. Newman-Keuls Test of Number of Class X Presentation Order, Error Rate 175

Table F21. Newman-Keuls Test of Number Of Items X Type, Error Rate 176

Table F22. Newman-Keuls Test of The Number Of Classes, Response Time 177

Experiment 2: Tables of Results (in Appendix H)

Table H1. ANOVA Summary Table for SWAT Rating 181

Table H2. ANOVA Summary Table for Error Rate 183

Table H3. ANOVA Summary Table for Response Time 185

Table H4. Newman-Keuls Test of Number Of Classes, SWAT Rating 187

Table H5. Newman-Keuls Test of Number Of Items, SWAT Rating 188

Table H6. Newman-Keuls Test of Number Of Classes X Number Of Items, SWAT Rating 189

Table H7. Newman-Keuls Test of Proximity, SWAT Rating 190

Table H8. Newman-Keuls Test of Presentation Order X Proximity, SWAT Rating 191

Table H9. Simple-Effect F-tests for Number Of Classes X Presentation Order, SWAT Rating 192

Table H10. Simple-Effect F-tests for Number Of Items X Presentation Order, SWAT Rating 193

Table H11. Simple-Effect F-tests for Number Of Items X Proximity, SWAT Rating 194

Table H12. Simple-Effect F-tests for Number Of Items X Presentation Order X Proximity, SWAT Rating 195

Table H13. Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 1, SWAT Rating 196

Table H14. Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 3, SWAT Rating 197

Table H15. Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 5, SWAT Rating 198

Table H16. Newman-Keuls Test Number Of Classes, Error Rate 199

Table H17. Newman-Keuls Test of Number Of Items, Error Rate 200

Table H18. Table H1.Newman-Keuls Test of Number Of Classes X Number Of Items, Error Rate 201

Table H19. Newman-Keuls Test of Number Of Classes, Response Time 202

Table H20. Newman-Keuls Test of Number Of Items, Response Time 203

INTRODUCTION

This dissertation addresses displays which provide a large volume of visual symbolic information relating the status of a dynamic system and/or environment. Systems which employ these complex information displays include air-traffic control, process control, CAD/CAM, and military Horizontal Situation Displays (HSDs).

When using complex information displays, operators often need to integrate, interpret, and act upon information rapidly and accurately. However, because of the volume of information and the rate at which it must be processed, it is possible that operators will become inundated and overloaded. Thus, the operator can be a limiting factor in overall system performance (after Goodstein, 1981; Kuperman and Wilson, 1986; Reinhart, Glynn, Dye, Takahama, and Snyder, 1988).

The current dissertation work involves a cognitive processing approach to understanding and reducing operators' cognitive workload. Since highly automated system interfaces require completion of cognitive processing tasks, workload can be described in terms of demands to perform cognitive tasks. Those cognitive tasks which exploit processing of Short-Term Memory (STM) stores are of primary interest in this work. That is, the current research effort sought to develop and demonstrate methods of reducing operator workload by reducing STM demands.

This research was completed as a series of two experiments. Experiment 1 assessed the cognitive workload required to categorize symbols shown on multi-symbol displays into task appropriate chunks. The experimental factors included presentation order (i.e., parallel versus Chunks-In-Sequence (CIS) presentation), locational proximity, and conceptual relatedness. Experiment 2 assessed the cognitive workload required to parse these object chunks into their individual component symbols. The factors considered in

Experiment 2 included presentation order, locational proximity, and conceptual relatedness. It was hypothesized that workload would be reduced to the extent that multi-item arrangements were pre-organized (by Artificial Intelligence (AI) systems) and presented to the operator in the form of task relevant chunks of information.

BACKGROUND

STM Scanning Rate

Sternberg (1966) demonstrated that a finite period of time is required to scan each item held in STM. On each trial, participants were presented one to six digits in a serial fashion. The digits presented are referred to as a Memory Set (MSET) of size n , where n varies between one and six. Presumably, participants stored these digits in STM. Then, a probe (i.e., test) digit was presented and the participant responded by pulling one of two levers to indicate whether or not the probe digit was a member of the MSET.

Sternberg demonstrated that response latency increases linearly at a rate of about 38 ms per item. Due to the linear nature of these findings, Sternberg concluded that participants scanned STM in a serial manner (i.e., one item after another, rather than scanning all items in parallel).

Cavanagh (1972) reviewed previously published studies that measured either *perceptual span* (the number of items a participant could hold in STM) or *scanning rate* (as measured by the Sternberg technique). Cavanagh concluded that, for any stimulus encoding class, there was an inverse linear relationship between perceptual span and scanning rate. Stated more simply, *in those encoding categories where participants could remember more items, they required less time to scan each item.* For example, in studies using digits as the encoding class, participants had an average memory span of 7.3 digits and required an average of 33.4 ms to scan each item. Across general studies using random forms as the stimulus class, participants had an average memory span of 3.8 forms and required an average of 68.0 ms to scan each item.

Two points from this discussion are important to designers of complex information displays. First, operators require time to process items held in STM. Second, some

cognitive subtasks (e.g., those requiring the scanning of more MSET items or MSET items of certain encoding classes) will take longer than others.

STM and Cognitive Workload

In a prior investigation related to the current research, Reinhart et al. (1988) demonstrated the role of STM processes in operator cognitive workload. Their participants performed a primary tracking task and a Sternberg memory scanning secondary task. Performance measures included response time on the Sternberg task and two subjective measures of workload: the Subjective Workload Assessment Technique (SWAT) and the Modified Cooper-Harper Scale. Results showed that as the number of MSET items increased, response times and subjective workload ratings increased accordingly. At least three conclusions can be derived from this work. First, a large portion of what is referred to as "cognitive workload" is concentrated in STM processes. Second, STM loading can be measured by traditional STM metrics (e.g., response time) as well as by cognitive workload measures (e.g., SWAT). Third, it is possible to measure STM demands associated with various modes of information presentation.

The work of Reinhart et al. married the study of STM to the study of cognitive workload. Thus, it seems reasonable to explore whether phenomena found to improve performance in STM tasks might also be applied to reduce workload in the context of complex system (e.g., crew member) display tasks. In this vein, the following literature review examines phenomena related to STM performance. One method of STM enhancement which has been especially fruitful in increasing STM capacity is the principle of chunking. Therefore, this principle serves to focus the literature review discussion.

CHUNKING

This section provides a definition of "chunking" and discusses issues related to chunking in visual information displays.

In his now-classic paper, Miller (1956) noted that most individuals can hold 7 ± 2 items of information in STM. This span of memory is independent of the number of bits of information held in each item. (The term bit refers to the amount of information required to resolve a binary choice.) One strategy people use to increase STM capacity is chunking. Chunking refers to the grouping of related elements held in STM. Since the number of chunks held is relatively fixed (i.e., 7 ± 2), yet the number of bits per chunk can vary, the amount of stored information increases by increasing the bits per chunk. For example, the four digits 1, 4, 9, and 2 can be grouped into a single chunk referring to the year in which Columbus discovered America. Theoretically, the single chunk "1492" occupies one slot in STM, thus freeing three slots to store additional items. One caveat, however, is that as item pronunciation time increases, perceptual span may decrease (Ellis and Hennelly, 1980).

Conceptually, the process of chunk formation might proceed as follows. An individual notices a relation among items in a set. The relation might be simple, such as some dots form a line, or complex, such as some military symbols form a specific threat. A label is assigned to the items comprising the relation. The individual items chunked under the label are stored in Long-Term Memory (LTM), but the label itself is stored in STM. To recall the items, the individual first recalls the STM label, then uses this label to prompt recall from LTM (after Chase and Simon, 1973). As shown in Figure 1, an individual's STM span may remain constant at seven items, but he can use chunking to increase the number

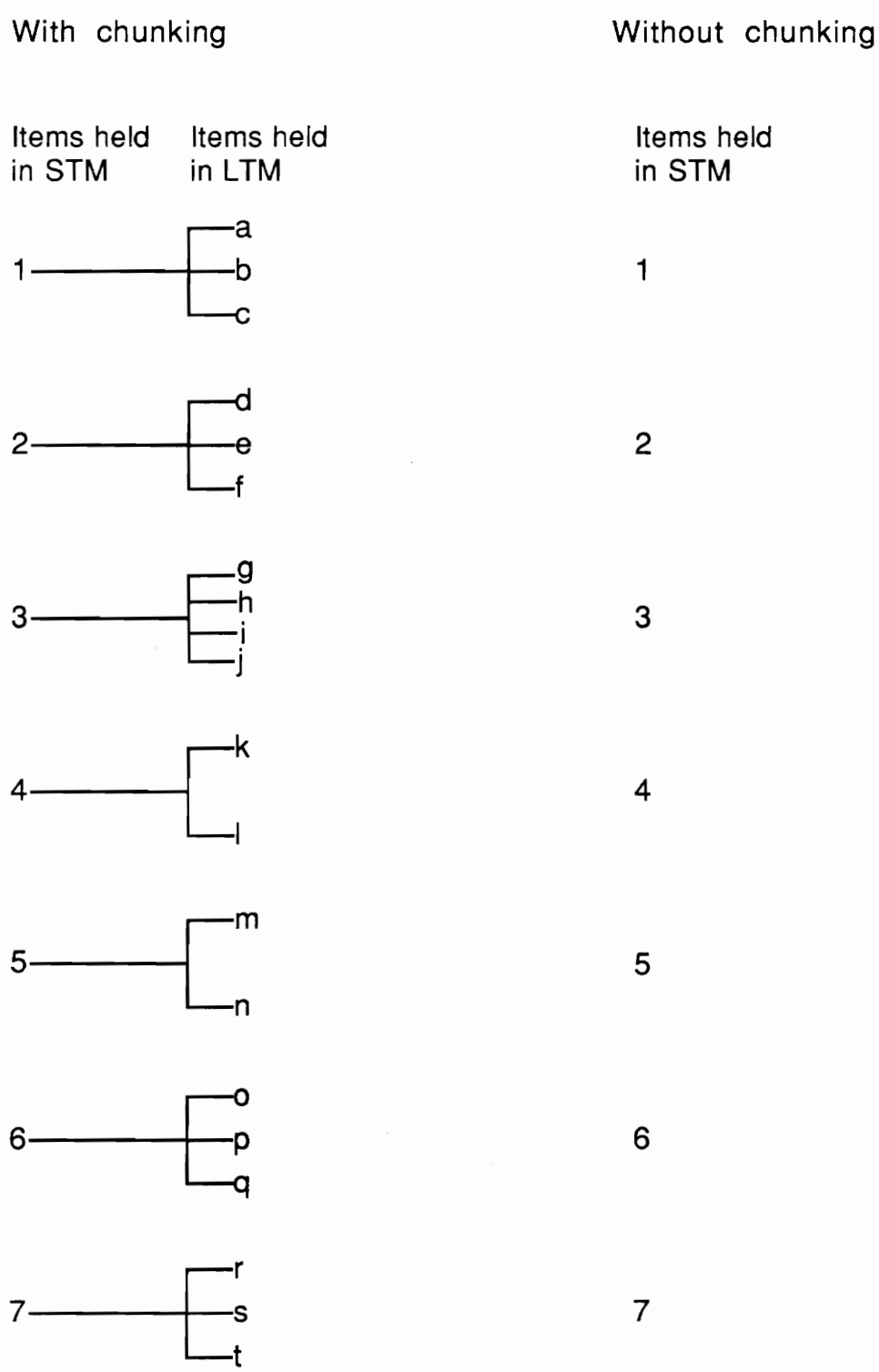


Figure 1. A conceptual model of chunking.

of items stored and recalled from memory stores. To recall an item embedded within a chunk, the person first recalls the STM label, then scans the items chunked under that label. The mechanics of this latter memory recall operation are termed chunk parsing.

Using a chess paradigm, Chase and Simon (1973) demonstrated that various strategies can be used to group information in STM. These experimenters used three chess-playing participants, a Master (M), a Class A Player (A), and a Beginner (B). Twenty partially played games were selected from chess books. Eight random chess games also were developed by perturbing actual game positions randomly on the board.

Two chess boards were used. A participant inspected the first board for five seconds. This board showed the partially played game. The participant then turned to the second board. This board was empty and chess pieces were placed along its side. The participant demonstrated recall by setting up the pieces on the second board in their correct locations. If the participant was unable to accurately recreate the chess game shown on the first board, the second board was cleared and the participant viewed the first board for an additional five seconds. This process was repeated until the participant accurately recreated the arrangement on the first board.

For trials in which there were 24 to 26 pieces on the board "M was able to place an average of about 16 pieces correctly on the first trial, while A and B placed about eight and four, respectively. M was able to reproduce the board perfectly in three or four trials, while A typically required about one or two more trials than M, but B took considerably more trials (as many as 14 in one case) " (Chase and Simon, 1973, p. 59). As shown in Figure 2, M's performance was inferior when pieces were randomly arranged. Regarding performance on the random game task, M stated he was troubled by positions that looked "unreasonable."

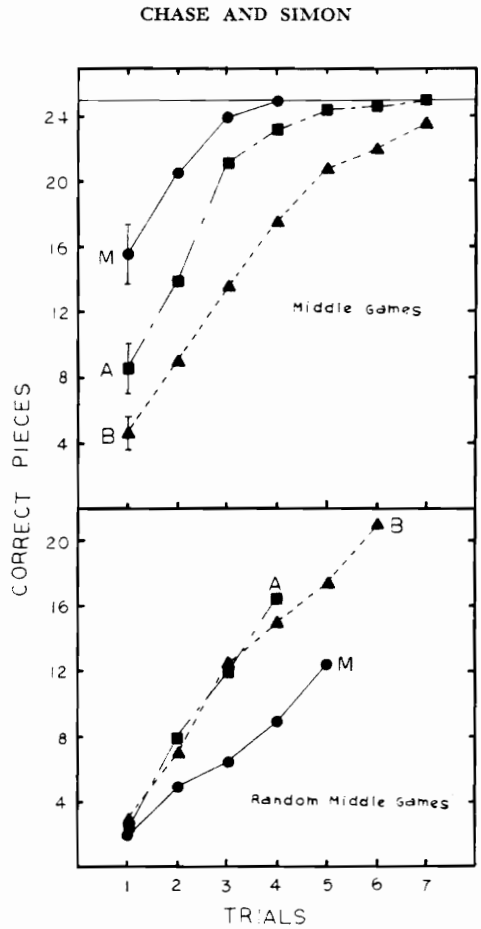


Figure 2. Learning curves of the master, class A Player, and beginner for the middle-game and random middle game positions. The brackets are standard errors on five positions (copied from Chase and Simon, 1973).

It seems M was constrained by normal STM limits. Therefore, M's rapid storage and phenomenal recall suggests that a chunking strategy was applied and, consequently, that chunking relations were more readily apparent to him.

In a continuation of this work, participants were asked to perform a second board re-creation task. In the second task, participants were allowed to look back-and-forth between two boards and, thus, they were not required to memorize an entire game. Chase and Simon described five relations that participants might use to chunk pieces: (1) *attack*, one chess piece or both chess pieces attacks the other, (2) *defense*, one or both of the pieces defends the other, (3) *common color*, e.g., both pieces are black, (4) *common piece*, e.g., both pieces are pawns, and (5) *proximity*, each piece stands on a square adjacent to the other.

The experimenters determined which pieces had been chunked by assessing between-piece placement times. To greatly simplify the procedure, a participant might place three pieces with between-piece intervals of 0.5 second, pause 4.0 seconds, then place three more with 0.5 second between-piece intervals. Here, Chase and Simon would have deduced two chunks, with three pieces each, had been created.

Proximity alone and in combination with other relations was the most frequently used chunking relation. While the results are difficult to interpret, participants may have used defense relations next in frequency, followed by attack relations. Same color and same piece relations were not used often (except when pieces also were related by proximity and/or defense). Pieces unrelated in all five ways were rarely, if ever, chunked.

Conceptual Versus Positional Chunking

The term *conceptual chunking* refers to item grouping in STM based upon a cognitive relationship. For example, Chase and Simon's attack and defense relations represent conceptual relationships, since pieces were related in terms of their function or meaning.

The term *positional chunking* refers to item grouping in STM based upon their perceived spatial positions. Chase and Simon's proximity relation falls within this category, but other spatial relations also apply. The Gestalt psychologist Wertheimer (1923; cited by Solso, 1988) suggested that some patterns of stimuli are organized naturally (or "spontaneously") into meaningful groupings. For example, we will perceive a house, not a discrete number of separate hues and brightnesses that constitute the image. Wertheimer suggested the following organizing principles:

Proximity. Elements close in time or space tend to be perceived together. For example, the lines in Figure 3a tend to be seen as three *pairs* of lines.

Similarity. Like elements tend to be seen together (Figure 3b).

Direction. Elements that comprise a continuous smooth direction tend to be seen together (Figure 3c).

Objective set. An organization perceived in one instance tends to be seen in immediately subsequent and similar cases. As shown in the right-hand arrangement in Figure 3d, we see dot pairs, even though no such grouping exists.

Common fate. Elements shifted in a similar manner from a larger set tend to be grouped together (Figure 3e).

Pragnanz. Figures will be made "good" or simplified. For example the gaps in the circle of Figure 3f will be ignored.

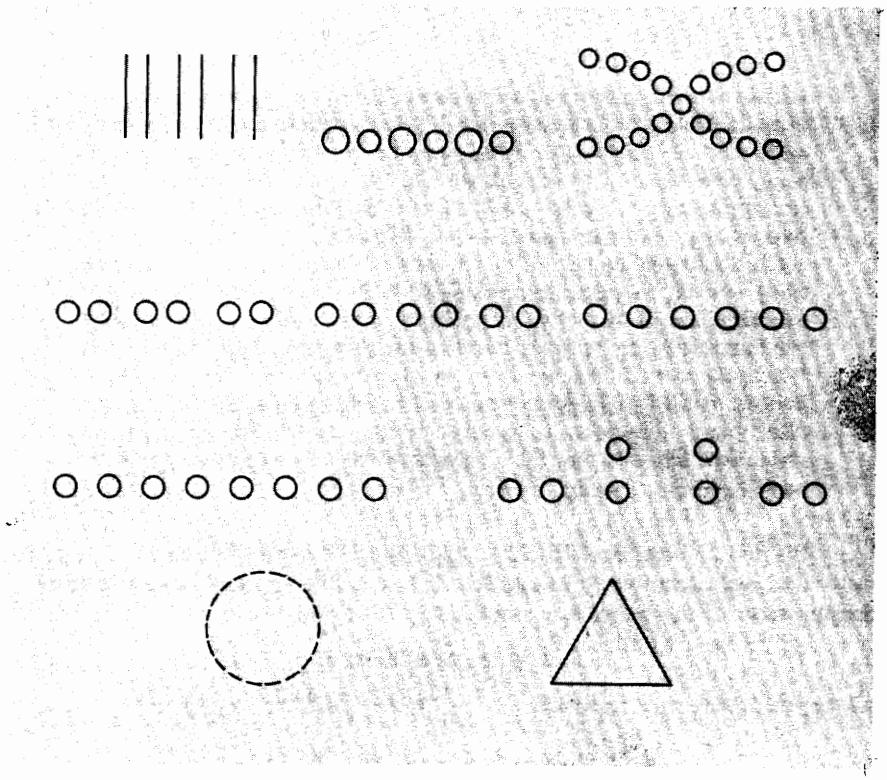


Figure 3. Examples of perceptual factors postulated in Gestalt psychology (copied from Solso, 1988, p. 56).

The processes of Gestalt form-making and STM chunking are not identical. Form-making is a perceptual process used to create mental images from visual sensations. Chunking is an STM process used to store and manipulate these mental images. Still, it seems likely that items perceived as comprising a form will be stored as a single chunk. If this chunk can be used to solve the problem at hand, the ideal operator will not reorganize the chunk. However, if the problem at hand requires a different organization, then the ideal operator will alter the chunk. The reorganization requires additional processing effort and, presumably, is associated with an increase in processing time. The implication of these Gestalt principles is that operator workload should be reduced by displayed information that allows processing of perceived forms (i.e., initial STM chunks) without cognitive reformatting or restructuring.

For example, Chase and Simon's participants made great use of spatial proximity or, perhaps, the Gestalt relations of proximity and direction. However, suppose Chase and Simon had asked participants to perform a different task, that is, selecting a next move for white. Here, chunks formed solely on the basis of proximity might be inappropriate. Solving this problem might be facilitated by a chess board design which aided Gestalt pattern recognition of the conceptual chunks. For example, all defense and attack relations might be shown as color coded lines. Later in this dissertation, it will be shown how grouping strategies might be applied to problem solving on multi-symbol displays.

Concluding this section, whether playing chess or using a complex information display, conceptual, as well as locational, relationships should be stressed. (Recall the operator must make decisions by using the display, not by memorizing it.) It therefore seems profitable to apply decision aiding and chunking methods that will facilitate conceptual chunking.

Chunking and Response Latency

It was suggested above (Figure 1) that chunking involves the formation of a two tier "data base." Each STM label is used to access a number of items stored in LTM. It has been demonstrated that chunking increases the number of items stored and recalled from memory. In this section, the effect of chunking on operators' response latencies will be considered. Three issues will be discussed: encoding times, recall times, and theoretical implications.

Encoding times. It is reasonable to hypothesize that the creation and storage of chunks requires more processing time than does direct encoding of information into STM. This issue was investigated by Smith (1954; cited by Miller, 1956). Smith taught participants different schemes for recoding incoming binary digits. For example, using the 4:1 scheme, the binary digits 0, 1, 1, 1 would be recoded and stored as the single digit "7". To report back the digits, the participant would recall the digit "7", then apply the decoding procedure to generate the original binary digits. Using this coding scheme, the total number of chunks decreased from 32 to 6. Thus, participants should have been able to store and recall more binary digits (i.e., bits of information) while still storing the same number of chunks. Unfortunately, participants were not able to recall the number of digits suggested by their perceptual spans (e.g., a participant with a perceptual span of 6 was able to remember more than 6 items, but less than 6 x 4 items). Smith hypothesized that recoding had not been automatic and that participants had lost part of one incoming chunk while trying to remember the translation of the previous chunk.

To test this hypothesis, Smith extensively practiced the task of recoding incoming binary digits. Finally, he was able to demonstrate that he could receive 40 binary digits (well beyond his STM capacity), then repeat them back without error. Smith's findings are

consistent with those of Chase and Simon. Specifically, recoding (i.e., chunking) times are reduced as expertise at pattern recognition increases.

Recall latency. A second question related to chunking and response latency is whether the recall of an item from a chunk (i.e., chunk parsing) will differ from recall of similar non-chunked items. For example, if a participant stored the digits 1, 4, 9, 2 as the number "1492"; how long would it take to parse this chunk and respond "yes" to the probe "4"? Sternberg (1954) showed that a finite time was required to scan each item held in STM, and that this time remained constant regardless of MSET size within the range of one to six items. It is not entirely known, however, exactly how people access chunked information. It is tenable that people scan STM labels until the correct one is found, then scan the items chunked under that label. In any event, it is likely that scanning a two-tier memory data base will be associated with a different latency per item than scanning a one-tier data base.

Cavanagh (1972). An interesting addendum to Cavanagh's (1972) conclusions is suggested. Cavanagh noted that there was an inverse linear relationship between perceptual span and scanning rate for any stimulus class. It is worthwhile to question the perceptual span/scanning rate relationships that develop when MSET sizes exceed STM capabilities. It can be expected that the number of items stored and recalled will increase as item scanning rate decreases.

Presentation Order

Presentation order describes the temporal spacing of information presentation. In parallel presentations, all MSET items are presented simultaneously. This order necessitates that MSET items be presented at different display locations. In serial presentations, MSET items are displayed sequentially. That is, each item is presented and

removed before the next item presentation. Sequential presentation permits displaying MSET items at a constant or at varying display locations. Figure 4 illustrates these two information presentation modes.

It is possible to predict the effects of parallel verses serial presentation order by considering the findings of Pollack (1952, 1953; as cited by Miller, 1956), who explored tonal pitch discrimination. Miller stated, "For example, if you can discriminate five high-pitched tones in one series and five low-pitched tones in another series, it is reasonable to expect that you could combine all ten into a single series and still tell them apart without error. When you try it, however, it does not work" (Miller, 1956, p. 84). Miller reasoned that the processing barrier does not exist at the perceptual level, since the participant has already demonstrated the ability to discriminate at least 10 tones. Instead, the barrier exists at the STM level. The participant must remember each unique tone and compare it with an incoming tone.

Miller noted that participants can discriminate numerous stimuli when making relative judgements (comparing currently displayed stimuli), yet can discriminate only a small number of stimuli when making absolute judgements (comparing a displayed stimulus to stimuli held in memory). He suggested that individuals are able to make larger numbers of relative judgements because of a reduced STM demand.

In regards to practical usage of these principles, information generally is presented in parallel. Observers of displays are able to view all symbols simultaneously and to create chunks. For example, four graphic icons might be perceived as a diagonal line and therefore, chunked under the STM label "diagonal line".

If information were presented serially, observers would need to make judgements between previously displayed items and currently displayed items in order to form chunks. This would increase the STM load in three ways. First, the observers might store and

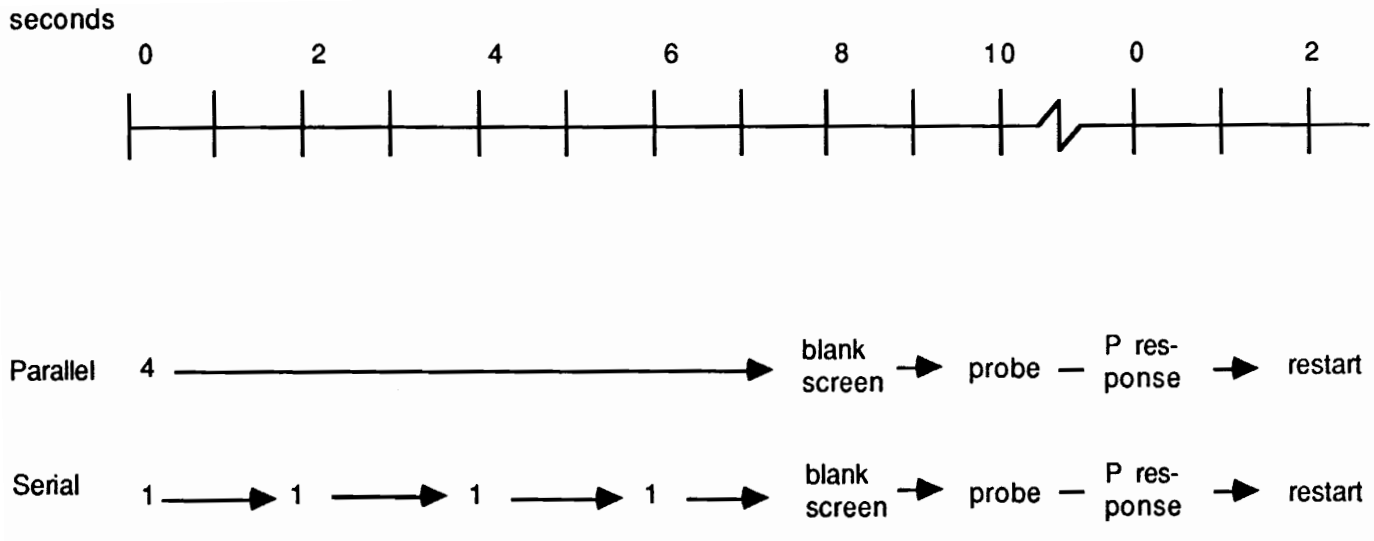


Figure 4. Parallel versus Serial Presentation, MSET size equals 4.

recall irrelevant items, thus consuming STM capacity. Second, to create chunks, observers would scan stored items, further loading the STM. Third, if observers were to create positional chunks they would need to store item locations, again stressing STM capacity. Conversely, with parallel presentation, positional chunks can be created at an iconic level.

Application of Chunking Principles to Visual Displays

This section provides a description of chunking concepts that might be incorporated into information display designs.

Goodstein (1981) stated that color and/or graphics visual displays present information in numerous ways. He noted these systems have been used too often in a *one measurement - one indication* approach. That is, each portion of the display provides one unit of information, just as a series of conventional displays might provide one unit of information per display. Therefore, operators must perform sequential observations and data processing to integrate these individual information units. This integration requirement loads STM and results in selection of a relatively restricted set of indicators.

Goodstein further suggested that operators solve problems at different "levels of abstraction" (e.g., global verses detail). Thus, a control display system should support decisions at various levels of abstraction. When an operator works at a high level of abstraction, the system might display global pieces of information. If the operator requires specific details (i.e., a "lower" level of abstraction) the system should "zoom in" on specific information components.

Bartram (1980) provided empirical support for the notion that some display formats are better suited to solving some problems. Bartram displayed hypothetical bus routes in one of four ways: (1) alphabetic list of stops, (2) ordered list of stops, (3) conventional road map, and (4) schematic map modeled after the London Underground map. Participants

were required to work quickly and to choose the route that required the fewest transfers. Individuals using the schematic map chose routes most quickly, followed by those who used the conventional map. The superiority of map formats over lists increased with task difficulty. Bartram added that, although it was not an *a priori* requirement, participants who used maps tended to choose routes requiring fewer stops.

Bartram suggested participants performed best with schematic maps because this format provided information requiring the least recoding. While the schematic map was best for the specific experimental task, it might not be appropriate for another task. For instance, a subway map is a good candidate for schematization because stops are well marked. A bus map might be a poor candidate for schematization if (1) stops are not well marked and a person might want additional location cues, or (2) a person might want to determine the shortest road travel distance. Also a less schematized map would be useful if the person wanted to determine which stop is closest to a certain address (Bartram, 1980). Applying Goodstein's levels of abstraction, we might suggest the use of a global schematic map. The traveler could then call up conventional maps (as well as information such as route times) as suited his purpose.

Samet, Geiselman, and Landee (1980) suggested there are numerous cognitive subtasks that must be completed when using military maps (see Table 1). In Table 1, it is apparent that cognitive subtasks associated with *symbology acquisition* and *processing single symbols* should be associated with encoding strategies used to group multiple symbols. For example, a grouping strategy should aid *learning*, *detection,* and *discrimination* of members of that group.

TABLE 1

Cognitive Subtasks Completed when using military maps

Symbology Acquisition

- o Perceptual learning- Acquiring a memory of a symbol's physical characteristics necessary for future recognition of the form.
- o Association - Acquiring a mental link between a symbol and the tactical unit or operation it represents.

Processing Individual Symbols

- o Detection - Realizing that a symbol is present on the display
- o Identification - Determining the meaning of a symbol.
- o Search - Determining the display location that contains a particular type of an identified form.
- o Tracking - Sustained detection of a moving symbol.
- o Updating - Acknowledging a change in symbol location or content

Processing Multiple Symbols

- o Comparison - Acknowledging sameness and/or differences between one displayed symbol and others.
- o Counting - Keeping track of the number of instances that a given form is encountered.
- o Pattern recognition - Interpreting the spatial arrangement of two or more identified forms.
- o Integration - Analytical assessment of the overall threat value of units in certain sectors of a display based upon multiple symbol features (including unit type, pattern, and relative distances).

Display Design Incorporating Chunking

This section initiates development of a visual display design that incorporates the chunking principles previously discussed. Design assumptions are as follows: First, personnel will have to perform various subtasks. Sample subtasks have been described in Table 1. Second, information can be presented in chunks such that recoding by the operator is minimized and task completion is accelerated. Third, it is likely that different decisions will be supported by different display formats. It is likely that some display formats need to support molar decisions (e.g., should I fly east and then north or north and then east) while others support molecular decisions (e.g., what is the best way to combat a specific airplane).

Efficiency and appropriateness. Based upon the previous literature review, it is clear that grouping displayed items aids the observer in creating a corporate entity for efficient memory, storage, and retrieval. We now consider the concept of appropriateness of the created information chunk. This is an important engineering concern since efficient display design must help the operator form chunks which are "correct" for the specific task at hand. For example, using the digits 2, 9, 4, 1, an operator might group the digits in pairs. Thus, the operator might remember 29 and 41. While efficiency of storage and recall has been enhanced, he is still no better off if the task will require him to report the year in which Christopher Columbus discovered America.

Another example of chunk appropriateness can be demonstrated with the children's game of jacks. Here a child drops a ball, picks up a specified number of jacks (without touching any others), then catches the ball before it hits the ground a second time. During different parts of the game, jacks are picked up two at a time, three at a time, etc. Suppose the jacks lay such that, due to proximity they form clusters of two, yet it is the child's turn to pick up three at a time. The child will process the information into chunks very

efficiently; the child will see sets of two. This is not an appropriate chunk for the task at hand, since the task requires picking up three jacks at a time.

The examples above demonstrate that, when displaying chunked information, the display format designer must consider the tasks to be performed. One useful approach is to view each subtask as a problem to be solved. Charness (1985) suggested that problems can occur in unfamiliar or familiar domains. Problem solving in unfamiliar domains requires the person to discover (induce) "operators" or rules describing connections among relevant concepts. Problem solving in familiar domains (e.g., chess or bridge, to those skilled in these games) relies on the use of extant rules. Rules can be derived from the set of knowable patterns. Regarding chess piece locations, de Groot (1966) sampled 192 master level games and found the black king occupied a specific board square in 125 of these games (after 21 moves). Similarly, Charness suggested moves are somewhat stereotypical. One such example is, "if you see an open file [column with no pieces on it] then consider putting a rook [castle] on it" (Charness 1985, p. 245). In fact, Charness suggested skilled chess players might out-pattern-recognize their opponents, rather than out-calculating them. Charness stated that developing a high level of expertise in a semantically rich field generally requires about 10 years.

RESEARCH OBJECTIVES

It seems likely that chunking items that are related proximally may be faster or easier than chunking items that are not proximal, but are related conceptually. Yet, operators of complex systems often rely on conceptual relationships. Experiments 1 and 2 below introduce and assess two novel display grouping formats that could be incorporated into electronic displays. The first method, called Tight Proximity, relocates symbols in the spatial dimension so that conceptually related symbols are spatially close. The second method, called Chunks-In-Sequence (CIS), groups related symbols in the temporal dimension. That is, symbols belonging to one class are presented simultaneously; they will blank out; then symbols belonging to a second class will be presented.

Experiments 1 and 2 explore the degree to which physical display parameters can facilitate STM processing, especially in the use of STM chunks. Therefore, each physical parameter is tied to a cognitive counterpart. For example, physical display complexity is manipulated in terms of the Number Of Classes and Number Of Items per class presented on the display screen. The cognitive counterpart of display complexity is the cognitive difficulty of a trial problem. The Number Of Classes relates the number of chunks which the participant creates while the Number Of Items relates the number of symbols the participant subsumes under each chunk. For example, if two classes of three items each are presented, participants may form three chunks of two items each in STM. Theoretically, participants might also chunk symbols belonging to different classes (e.g., subsume *airplane*, *dog*, and *fork* under one chunk label).

Display grouping strategies describe the way information is physically associated on the display screen. In the experiments below, display grouping is in terms of the variables Presentation Order and Spatial Proximity. Some levels of display grouping provide

physical relations among displayed symbols (i.e., Tight Spatial Proximity, CIS Presentation Order). The cognitive counterpart of display grouping is "chunkability" of the information. It is hypothesized that display formats which provide increased grouping allow participants to encode more efficient and appropriate chunks. Chunking, in turn, allows enhanced completion of the experimental tasks.

Experiment 1 requires participants to categorize information into meaningful chunks and provide labels for these chunks. Experiment 2 requires participants to identify and recall large numbers of displayed items. Thus, the latter experiment requires observers to organize information into chunks, store and recall the chunks, and then parse those chunks into component items. It is hypothesized that display formats which group chunkable items will facilitate efficient and appropriate encoding.

EXPERIMENT 1: METHOD

Experiment 1 required participants to search a display containing numerous object symbols or icons. Participants were asked to categorize the symbols under chunk labels and then store these labels in STM. It was hypothesized that display formatting methods which grouped same-class symbols (i.e., Tight Proximity, CIS order) would help participants categorize the symbols. The efficiency and ease of symbol categorization was assessed by means of error rate, Subjective Workload Assessment Technique (SWAT), and response time measures.

Results will be interpreted in terms of the chunking principle. However, it is noted that it also might be profitable to interpret results in terms of concept formation literature. Concept formation relates to the means by which individuals determine whether an item should be categorized as an example of a concept (Medin and Smith, 1984).

Participants

Ten male, undergraduate students served as participants. Students were recruited through an advertisement in the school newspaper and received $25 for their participation. Participants were screened for contrast sensitivity with the Vistech Vision Contrast System (Vistech, 1986) test for near vision. The criterion was that corrected visual contrast sensitivity should fall within the top 90 percent of the population. All participants passed the selection test.

Apparatus, Stimuli, and Experimental Environment

A laboratory microcomputer (Model IBM PS/2 80) was used to control the experiment and collect all response data (i.e., response latencies, response error rates, and SWAT

ratings). A color CRT monitor (Model IBM 8514, 35-cm diagonal, 640 x 480 pixels) provided the video display. The icon set consisted of 64 symbols, equally divided into eight categories (see Table 2 and Appendix A). Maximum symbol dimensions were 0.9 cm high by 1.1 cm wide (1.13 by 1.38 degrees of visual angle assuming a 45.7-cm viewing distance), contained within an 18 by 30 pixel character cell.

A two-button mouse (Model IBM 6450350) served as the input device, and a high-level computer language (Microsoft C, Version 5.1) was used for screen formatting, response time measurement, and data collection and storage. The internal MS-DOS clock provided timing for program flow-control and response time measurements.

Task

A modified Sternberg paradigm (Sternberg, 1966) was employed for the experiment. On each trial, a total of 2 to 25 symbols were presented at various display locations. Participants were tasked to store the symbol class names (not the individual symbol names) in STM. For example, if the screen showed the symbols: *frog*, *elephant*, *airplane*, and *helicopter*, the participant was required to store the class names "animal" and "weapon".

Following the symbol display, the screen was blanked for 2 s and then a single probe symbol was presented at the center of the screen. The participant used the mouse buttons to respond "yes" (left button) or "no" (right button) to indicate whether an item from the probe class was presented during that trial. For example, given the symbols *frog*, *elephant*, *airplane*, and *helicopter* and the probe symbol *bear*, the correct response is "yes" (left button pressed and released), since *bear* belongs to the stimulus class "animal". Participants were instructed that response speed and accuracy were important.

TABLE 2

Symbol Set used in Experiments 1 and 2.

Class	Items
Letter	A, B, C, D, E, F, H, I
Digit	1, 2, 3, 4, 5, 6, 7, 9
Geometric Shape	triangle, square, rectangle, diamond, circle, oval, semi-circle, pentagon
Military Icon	airplane, helicopter, missile (air-to-air), tank, truck mounted missile, radar, gun, warship
Animal	fish, frog, turtle, snake, bird, dog, bear, elephant
Human Body Part	hair, ear, eye, nose, mustache, lips, hand, leg
Traffic Sign	Blacksburg Transit (local bus line) sign, stop sign, stop light, yield sign, speed limit sign, highway route marker, keep right sign, no-U-turn sign
Kitchen Item	fork, spoon, pot, electric mixer, cup, pitcher, breadboard, goblet

Design

A five-factor within-subjects factorial design was employed to define the experimental trial conditions (Figure 5). The five factors were: Number Of Classes (4 levels), Number Of Items per class (3 levels), Presentation Order (2 levels), Proximity (3 levels), and Probe Type (2 levels). Each participant viewed each of these 144 possible combinations in a unique random order during a block of trials. Each participant completed four blocks of trials, lasting about one hour each.

Number Of Classes (4 levels). This factor refers to the number of classes presented on a trial, and it was varied across two, three, four, or five classes. For example, if the Number Of Classes for a trial was set at two, then two classes (e.g., "animals" and "letters") were selected randomly from the eight available classes. This variable was provided so that Spatial Proximity and Presentation Order, the variables of greatest interest, could be examined across a full range of display complexity.

Number Of Items (3 levels). This factor refers to the number of items per stimulus class that were displayed during a trial, and it was varied across one, three, or five items. For a trial where the Number Of Classes was two (e.g., "animals" and "letters") and the Number Of Items per class was three, then six total items were displayed. The specific symbols shown during any trial were selected randomly. The Number Of Items, like the Number Of Classes, was included in the experimental design so that spatial Proximity and Presentation Order could be examined across a full range of display complexity.

Spatial Proximity (3 levels). Spatial Proximity refers to the spatial separation among items within each symbol class. At the "Random" (non-chunked) level, symbols were assigned randomly across the entire display surface. At the "Medium" level, each item within a stimulus class was assigned randomly within a 4.5-cm radius on the display screen. The origin of the radius was assigned to a random position within the usable

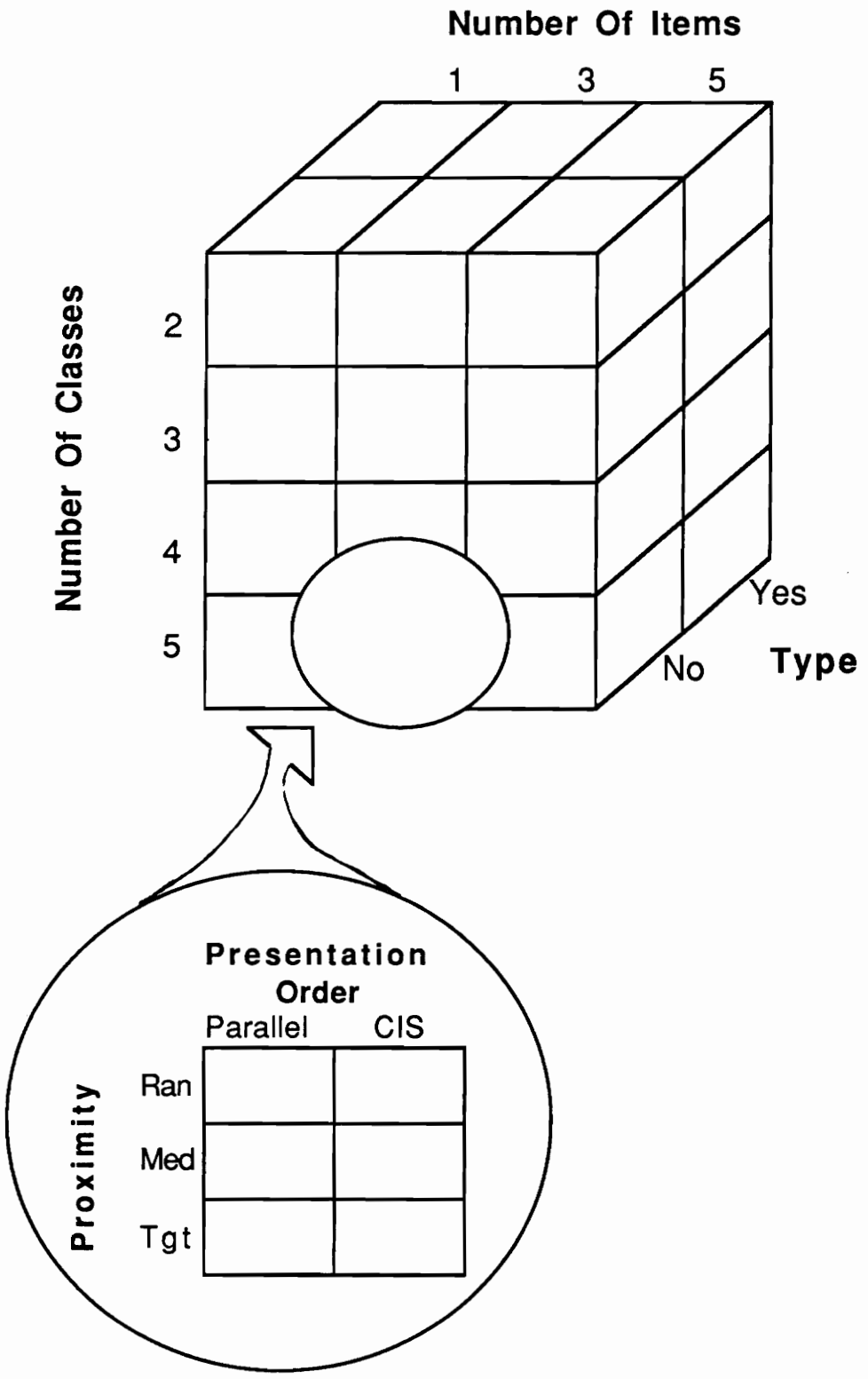


Figure 5. Graphical Representation of Independent Variables for Experiments 1 and 2.

display screen area. At the "Tight" level, each symbol was assigned randomly within a 1.5-cm radius and, again, the class center point was randomly positioned across the display screen.

Presentation Order (2 levels). Presentation Order refers to the temporal spacing of displayed symbols. In the "Parallel" presentation mode, all symbols (i.e., all items from all classes) were presented simultaneously (Figures 6 - 8). A time of 0.8 s was allocated for each symbol class, not each class item. For example, if two symbol classes (e.g., "animals" and "letters") were selected for presentation on a trial, the total presentation time was 2 classes x 0.8 s/class equals 1.6 s.

In the "Chunks-in-Sequence" (CIS) presentation mode, all symbols belonging to any one class were presented simultaneously, but symbol classes were presented sequentially (Figures 9 - 11). For example, if the Number Of Classes equaled 2, symbols belonging to one class (e.g., "animals") were presented for 0.8 s and then blanked out. Next, symbols belonging to the second class (e.g, "letters") were presented for 0.8 s and then blanked out.

Response Type (2 levels). The two levels of Type describe whether the correct answer for a trial was "yes" or "no". In the "yes" case, the probe symbol was a random symbol that had not been presented, but which belonged to one of the stimulus classes displayed on that trial. For example, if the symbols: *frog*, *elephant*, *airplane*, and *helicopter* were displayed, *bear* or *radar* would have been acceptable probes.

In the "no" case, the probe symbol was drawn randomly from a class that had not been presented during the trial. For example, if the symbols: *frog*, *elephant*, *airplane*, and *helicopter* were displayed during the presentation phase, *fork* or *leg* would have been acceptable probes.

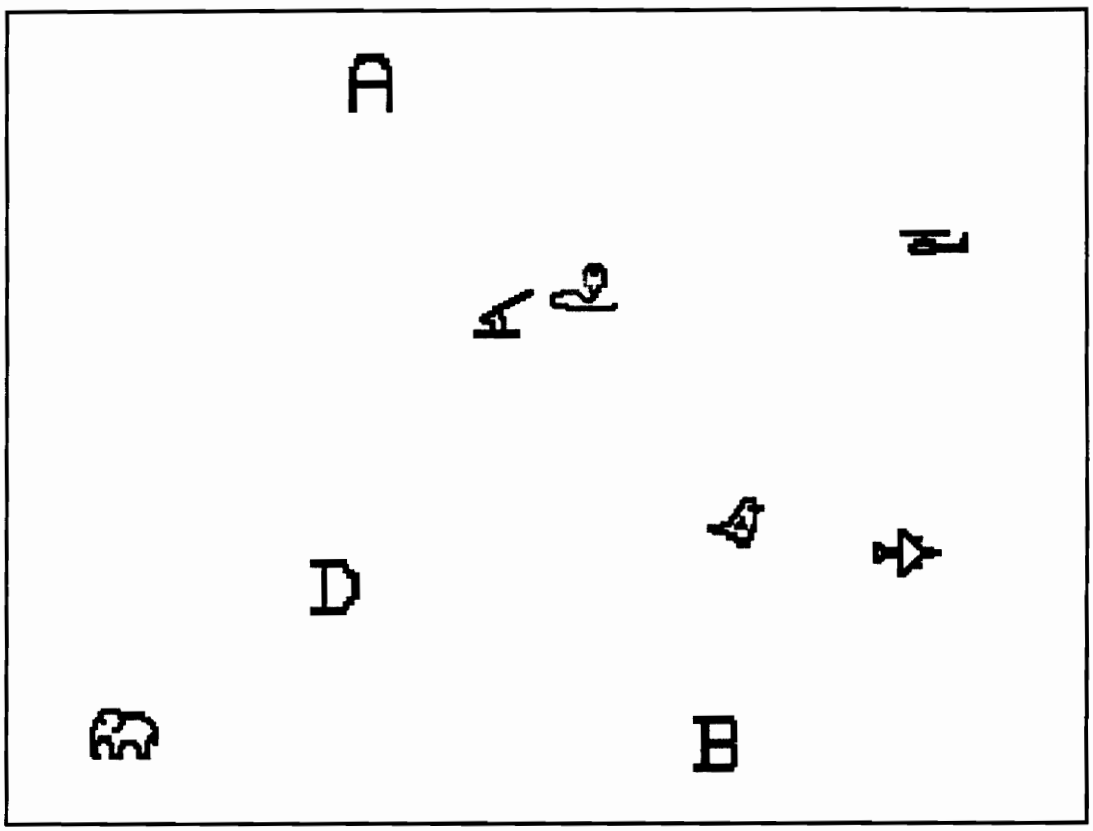


Figure 6. Display Showing Parallel-Presentation-Order/Random-Proximity. Display size shown is approximately 1:2 relative to the actual display. Symbol size is approximately 1:1, but has inferior integrity, relative to the actual display.

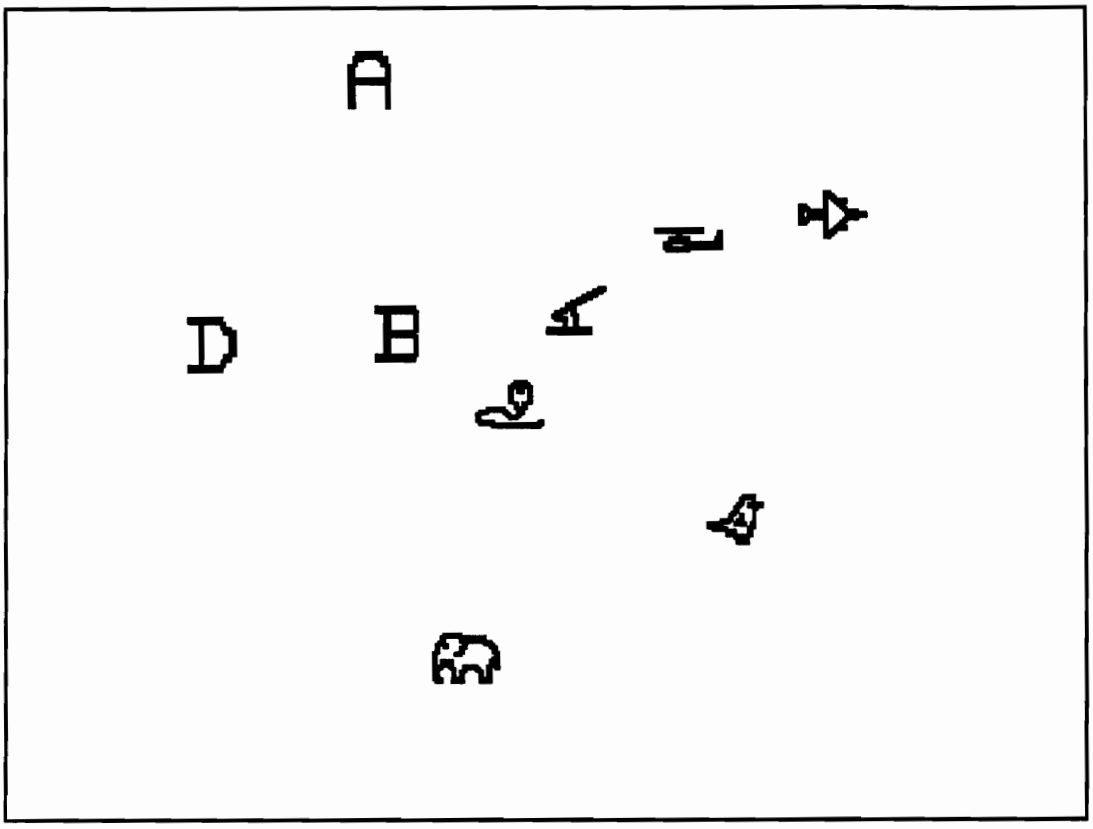


Figure 7. Display Showing Parallel-Presentation-Order/Medium-Proximity. Display size shown is approximately 1:2 relative to the actual display. Symbol size is approximately 1:1, but has inferior integrity, relative to the actual display.

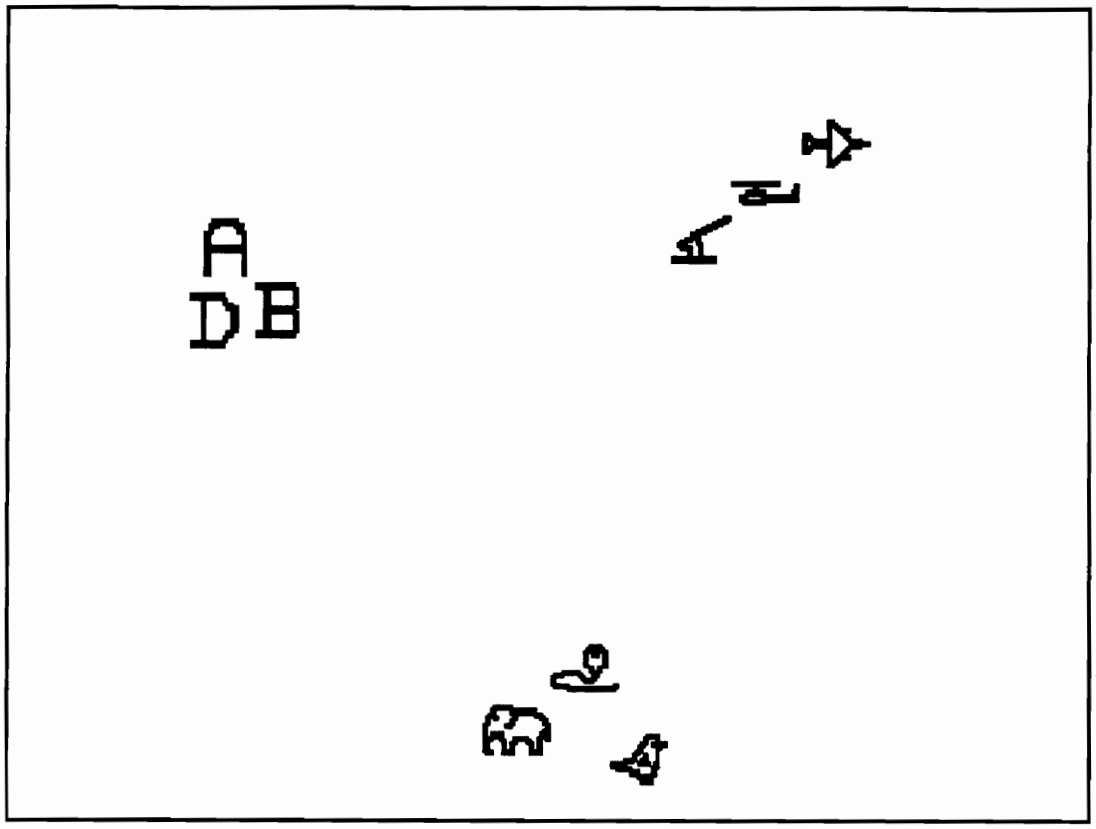


Figure 8. Display Showing Parallel-Presentation-Order/Tight-Proximity. Display size shown is approximately 1:2 relative to the actual display. Symbol size is approximately 1:1, but has inferior integrity, relative to the actual display.

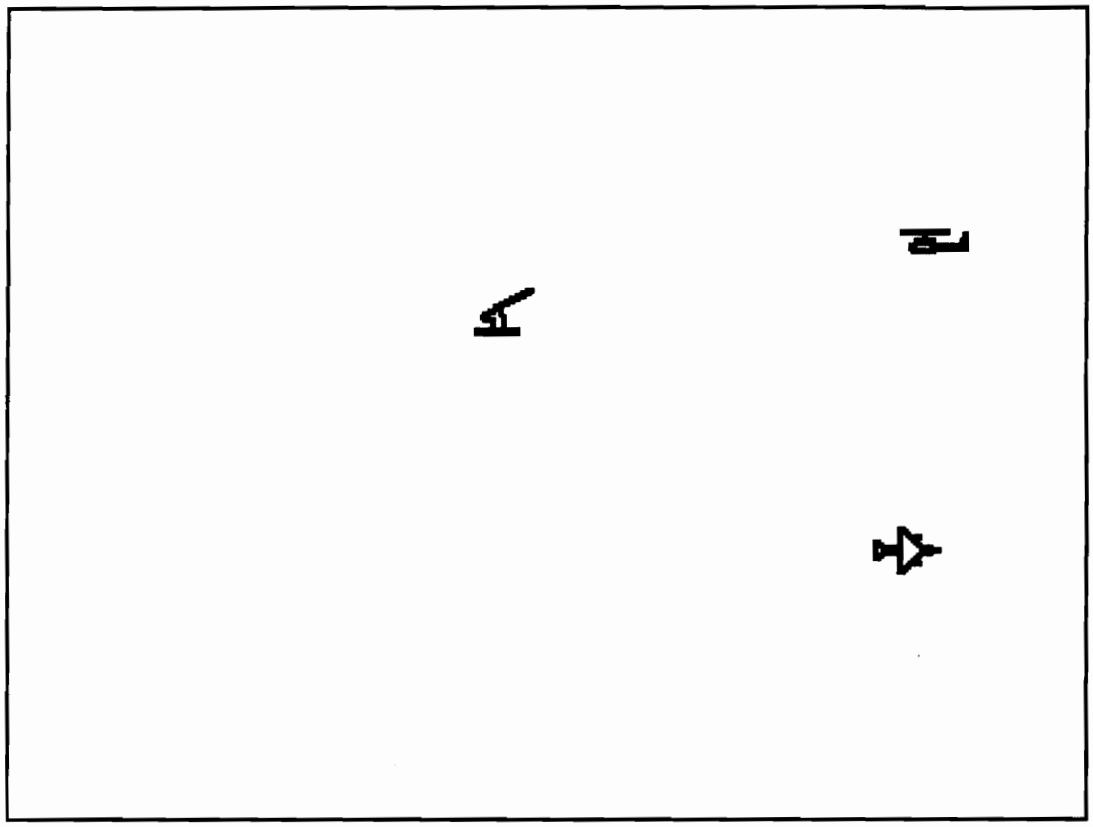


Figure 9. Display Showing CIS-Presentation-Order/Random-Proximity, first 0.8 s. Display size shown is approximately 1:2 relative to the actual display. Symbol size is approximately 1:1, but has inferior integrity, relative to the actual display.

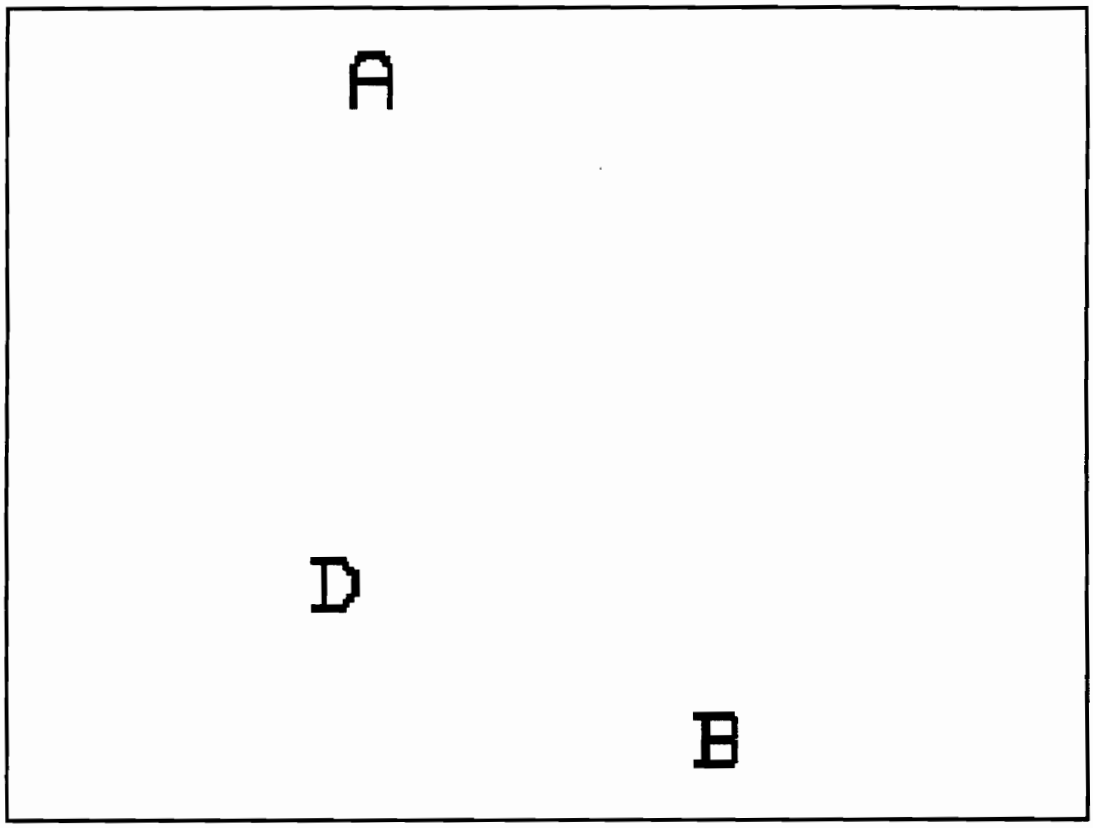


Figure 10. Display Showing CIS-Presentation-Order/Random-Proximity, second 0.8 s. Display size shown is approximately 1:2 relative to the actual display. Symbol size is approximately 1:1, but has inferior integrity, relative to the actual display.

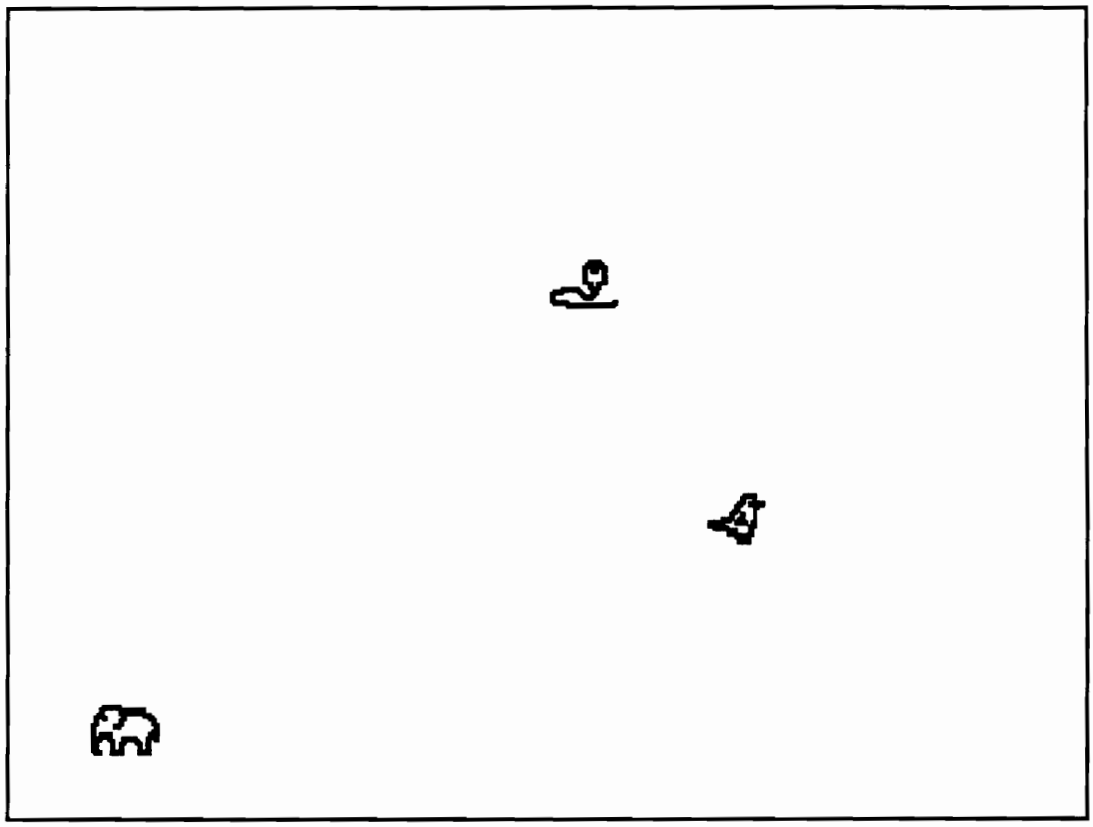


Figure 11. Display Showing CIS-Presentation-Order/Random-Proximity, third 0.8 s. Display size shown is approximately 1:2 relative to the actual display. Symbol size is approximately 1:1, but has inferior integrity, relative to the actual display.

Blocks (3 levels). Each participant completed four replications of 144 trials in the experiment. The first replication was not analyzed but served to familiarize participants with the task, symbol set, and SWAT rating system. Response data from blocks 2 through 4 were used in the statistical analysis.

Dependent Measures

Three dependent measures were collected during each trial. Response time was defined as the time (in ms) from probe onset to participant response (i.e., pressing the "yes" or "no" button). Error rate was defined as the percentage of trials in which participants made an incorrect response to the probe stimulus. Finally, SWAT ratings, describing a participant's opinion of cognitive workload, were collected for each trial (see Appendix C for a description of the SWAT metric).

Procedure

Individuals participated over a five-day period, providing a total time per individual of between 4 and 5 hours. On the first day, participants completed the informed consent form (see Appendix B), the vision test, and the SWAT card sort. (See Appendix C for an explanation of SWAT and Appendix D for SWAT instructions to participants.) Participants also learned the experimental task (see Appendix E for task instructions to participants) and completed one full practice replication. On each of the days 2 through 5, participants completed five practice trials followed by one full block of 144 trials.

EXPERIMENT 1: RESULTS AND DISCUSSION

Due to the large volume of data, it is useful to consider the process or strategy participants might have used to solve each problem. It is suggested that during a trial, a participant might :

- o glance at the *bear* symbol, classify it as an "animal" and store the "animal" class label in STM.
- o shift his glance to the *fork* symbol, classify it as a "kitchen item", scan the labels stored in STM to determine whether he already has a "kitchen item" label, decide "no", then store the "kitchen item" class label in STM.
- o shift his glance to the *dog* symbol, classify it as an "animal", scan the labels stored in STM to determine whether he already has an "animal" label, decide "yes" and <u>not</u> store the "animal" label a second time.
- o etc. (It is likely that in many instances the presentation set disappeared before the participant had time to view all of the symbols.)
- o After symbol presentation, the probe was presented and the participant scanned the list of class names stored in his STM.

Based on this cognitive scenario some points become evident. First, after the participant has stored a class label in his STM, all other symbols belonging to that class act as noise. Display formats that facilitated chunking were useful because they helped the participant choose which symbol to process next (i.e., helped choose a symbol from a class not already sampled).

Data Analysis Overview

For each dependent measure, there were 432 data observations (3 blocks x 144 trials per block) across the factorial combinations of Blocks (3 levels), Number Of Classes (4 levels), Number Of Items (3 levels), Presentation Order (2 levels), Proximity (2 levels), and Type (2 levels). Thus, the 10 participants provided a total of 4320 data points for each dependent measure. The data for each dependent measure (i.e., SWAT, reaction time, and error rate) were subjected to a five-factor, within-subjects Analysis of Variance (ANOVA) procedure. Data were averaged across blocks during the ANOVA. Complete tabulations of the ANOVA results are listed in Tables F1 - F3. The findings of each ANOVA are discussed below.

SWAT Ratings

Prior to data collection, a SWAT scale development procedure was completed by each participant. Each participant sorted 27 cards representing low to high workload situations. Scale values for each individual were calculated using the additive polynomial model incorporated in the SWAT conjoint analysis software (Reid, Potter, and Bressler, 1989). Scaling values provided "transformation functions" so that subsequent dimension ratings could be converted into overall integer ratings for each individual.

During scale development, there were no independence and six joint independence violations across the card sorts for all 10 participants. This relatively low number of violations fell well below the suggested maximum of 20 independence and 20 joint independence violations (Reid et al.,1989), and indicated that each participant's card sort was internally valid. Therefore, no participant was required to repeat the scale development procedure.

Kendall's W, the Coefficient of Concordance, across participants was 0.78. This value exceeded the suggested minimum of 0.75 (Reid et al., 1989) and indicated that participants provided homogeneous card sorts. Accordingly, a single scaling solution was used to represent their composite view of workload.

The previously developed scaling values were used to transform the three dimension ratings (i.e., *time, mental effort, stress*) from each trial into a single overall SWAT rating ranging between 0 and 100. A SWAT rating of 100 represented the highest possible workload. The average SWAT rating across all 4320 trials was 27.38.

Effects due to display complexity. Display complexity was controlled by manipulating the Number Of Classes and Number Of Items per class. As illustrated in Figure 12, SWAT ratings increase monotonically as the Number Of Classes increases ($p < 0.0001$). A Newman-Keuls test (Table F4) indicates that all levels of Number Of Classes differ significantly ($p < 0.05$). As shown in Figure 13, SWAT ratings also increase monotonically as the Number Of Items increases ($p < 0.0001$). A Newman-Keuls test (Table F5) indicates that all levels of Number Of Items differ significantly. Viewing Figures 12 and 13, it appears that workload increases more rapidly across the Number Of Classes than across the Number Of Items.

The Number Of Classes by Number Of Items two-factor interaction also is significant ($p < 0.0421$). As demonstrated in Figure 14, the effect of Number Of Classes at each level of the Number Of Items variable, appears reasonably constant. Therefore, this interaction does not modify the general findings based on separate main effects in a practically significant manner.

An auxiliary analysis was performed on the total number of symbols (i.e., Number Of Classes x Number Of Items) displayed. As shown in Figure 15, SWAT ratings increase

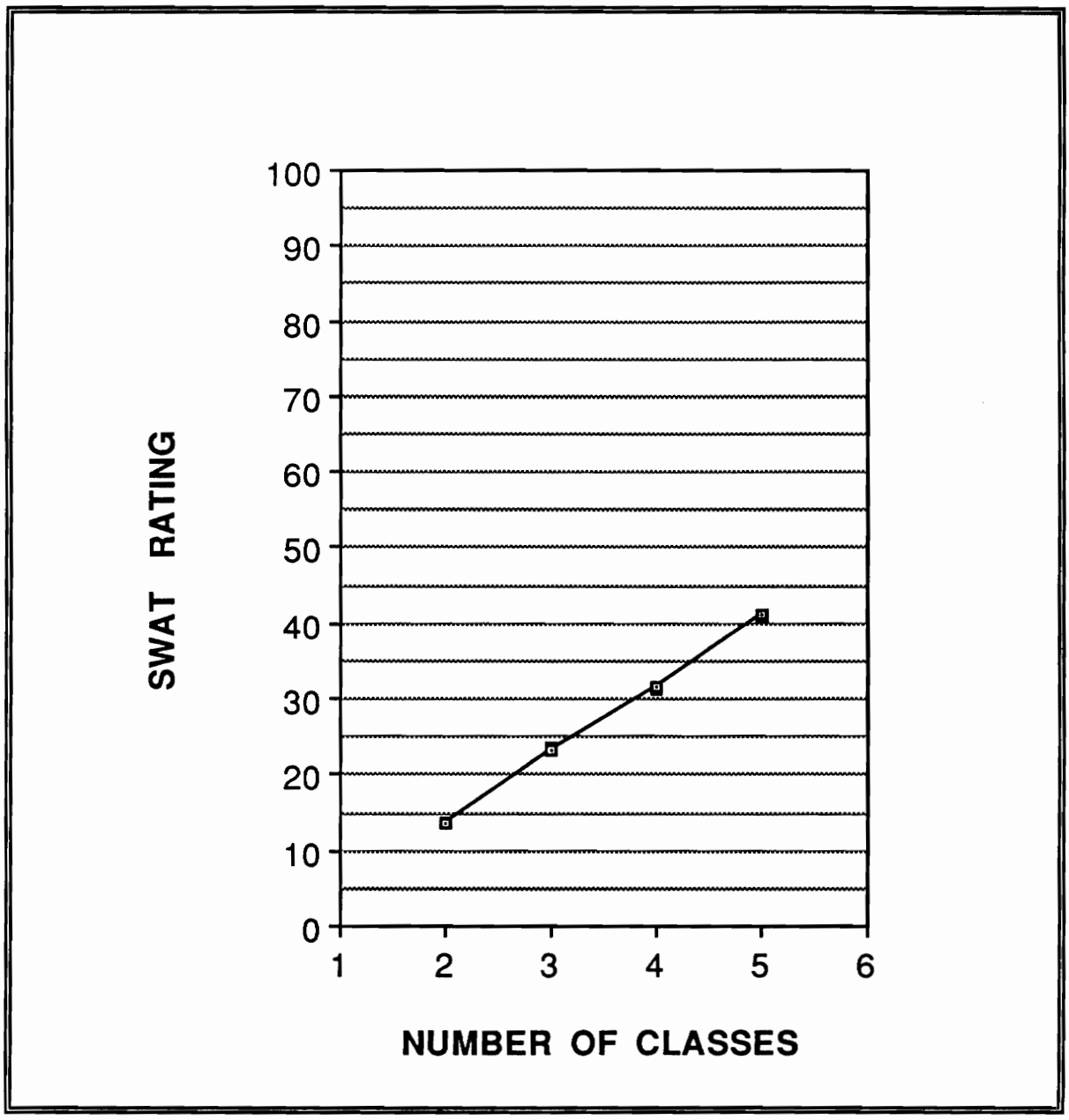


Figure 12. Effect of Number Of Classes, collapsed over subjects (N = 10), Blocks, Number Of Items, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars (not visible due to small size) indicate $\pm$ 1 standard error of the mean ($s_{\bar{x}}$).

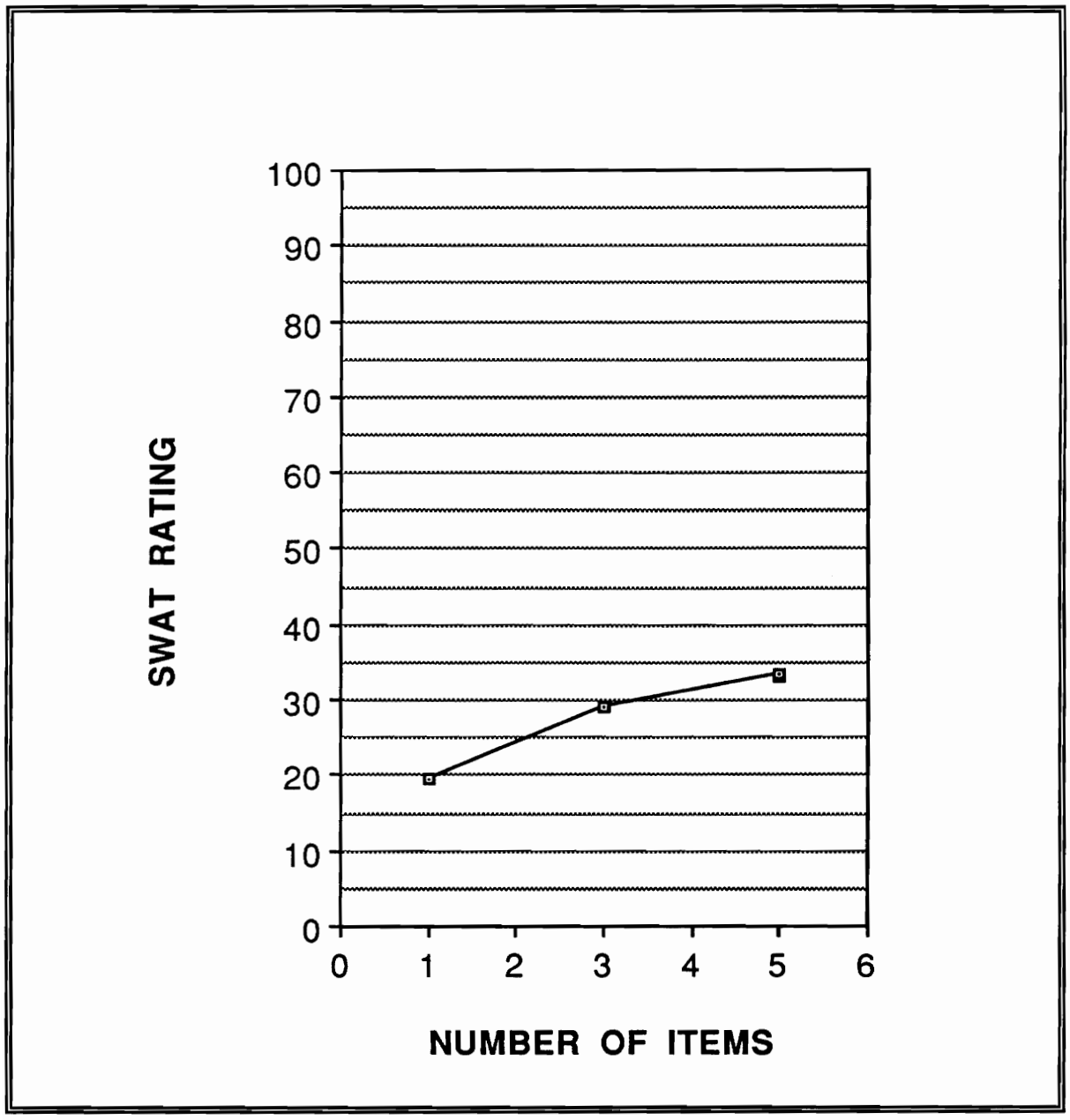


Figure 13. Effect of Number Of Items, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars (not visible due to small size) indicate $\pm 1\ s_{\bar{x}}$.

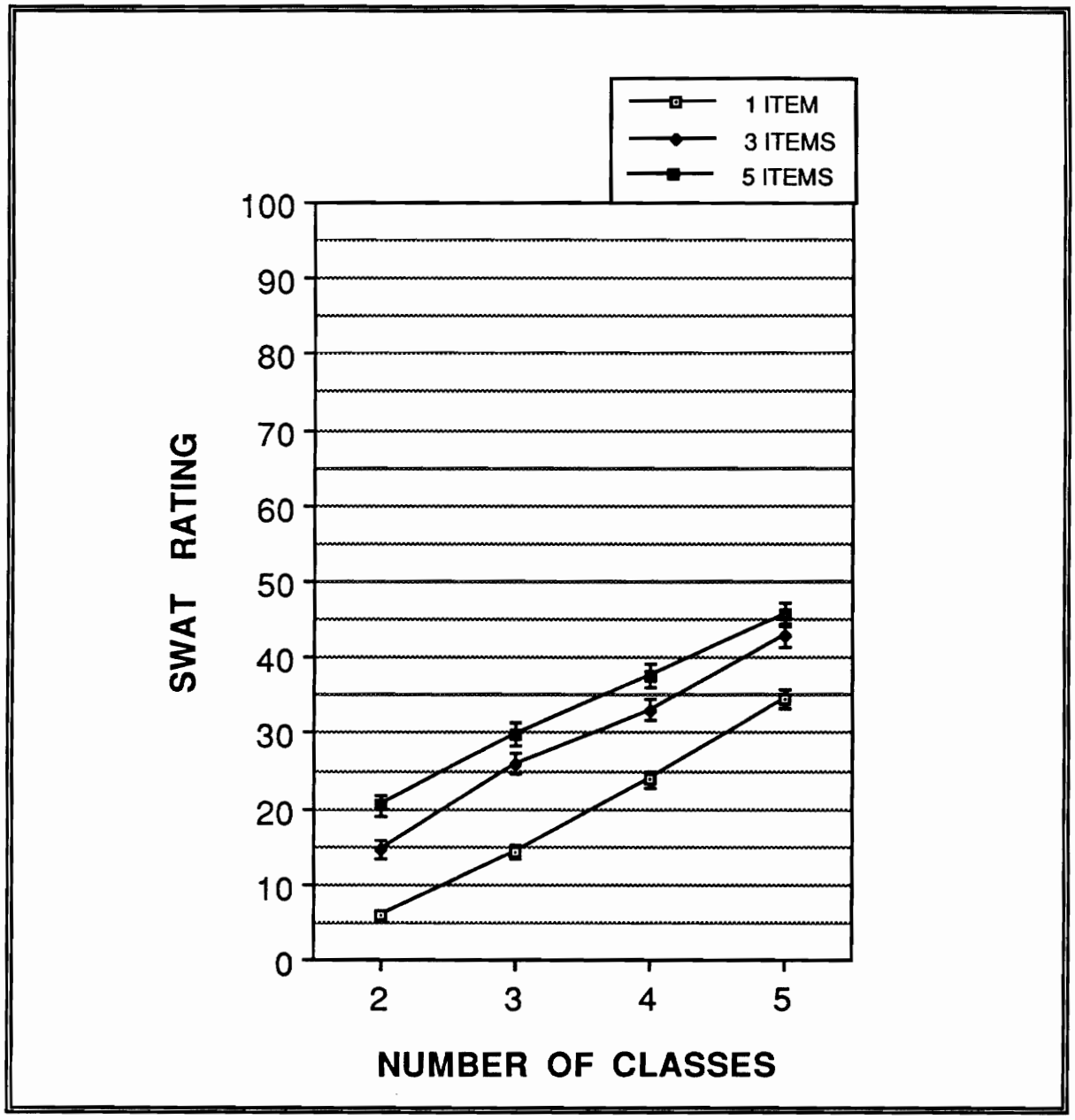


Figure 14. Two-factor interaction of Number Of Classes X Number Of Items, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 1 ($p < 0.0421$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

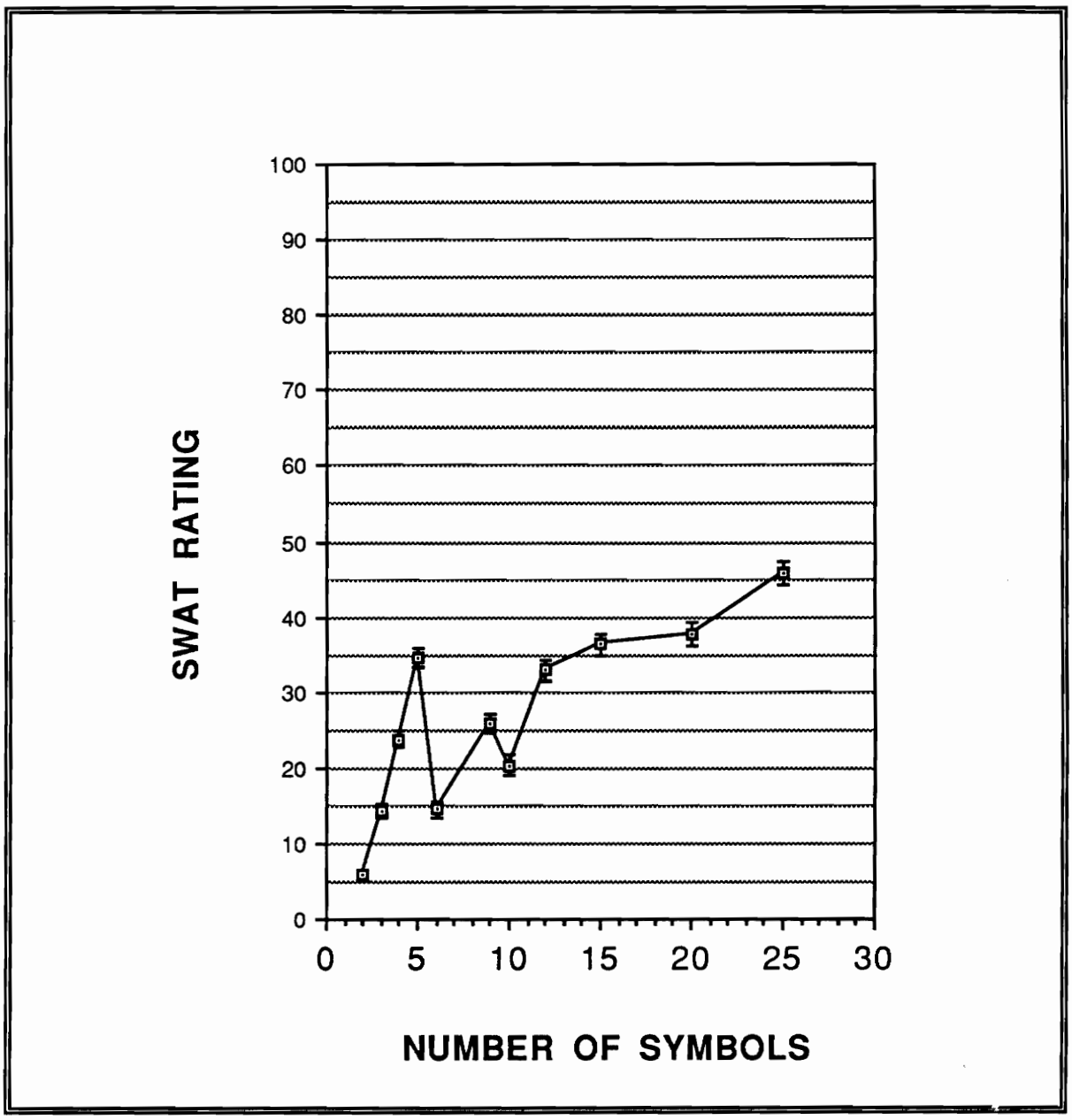


Figure 15. Effect of the number of symbols, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 1. Error bars indicate $\pm 1\ s_{\bar{x}}$. Note: There were two cases in which the number of symbols equalled 15. Average means and $s_{\bar{x}}$ s are shown.

linearly up to five symbols, then decrease sharply followed by a gradual increase up through 25 total symbols. These trends suggest that different processing mechanisms may underlie SWAT judgements; one mechanism operates up to five total symbols and a second mechanism operates when more than five symbols are presented. A Newman-Keuls test (Table F6) indicates that most of the SWAT rating averages depicted in Figure 15 differ significantly ($p < 0.05$).

It is interesting that Number Of Classes has a more marked effect on subjective workload than does Number Of Items. Number of Classes may increase task difficulty in three ways. First, it increases the size of the MSET list which participants had to store and iteratively compare against the class labels of incoming symbols. Second, it increases the number of displayed items. Finally, it increases the size of the MSET each participant had to scan following probe presentation. Conversely, the Number Of Items only increases the number of items which the participant had to sift through in order to extract relevant class names. That is, the Number Of Items provides a noise-to-symbol ratio.

It is also interesting to consider error rates on a per symbol basis. Figure 15 shows that subjective workload increases with the total number of symbols, but that this increase is not monotonic. It is suggested that the difficulty peak at 5 symbols occurs because there were five chunks (i.e., class labels) with one item per chunk. The participant was tasked to store this string, a moderately difficult problem after numerous trials, while still only accounting for five symbols. In the 25-symbol case, participants were still required to store five chunks, but these accounted for five symbols per chunk.

Effects due to display grouping. In this experiment, the effect of display grouping were manipulated by two variables: Presentation Order and Spatial Proximity. CIS Presentation Order provided symbol classes that were grouped through temporal serial presentation, while in Parallel Presentation Order symbols were not grouped temporally. Across the

Spatial Proximity levels, symbol classes were grouped through a spatial "clustering" technique. The Tight Proximity condition, for example, provided symbol classes that appeared as small clusters of similar items.

As hypothesized, display formats which grouped same-class symbols are associated with lower subjective workload ratings. The main effect of Presentation Order is significant ($p < 0.0006$). As shown in Figure 16, the average SWAT rating is 33.23 for the Parallel Presentation Order and 21.54 for the CIS Presentation Order. The main effect of Proximity also is significant ($p < 0.0001$), indicating monotonic decreases in workload as Spatial Proximity increases (Figure 17). As shown in Table F7, SWAT ratings differed at each level of Proximity ($p < 0.05$).

The two-factor interaction between Presentation Order and Proximity also is significant ($p < 0.0001$). In Figure 18 (and Table F8), it can be seen that Proximity does not affect SWAT ratings under the CIS Order, but that SWAT ratings decrease as Spatial Proximity increases under the Parallel Order condition. Note that the Parallel-Order/Tight-Proximity condition produces low SWAT ratings similar to the CIS presentations.

It is suggested that the Parallel-Order/Random-Proximity condition produces high workload since participants often re-sampled stimulus class (i.e., chunk) members. Redundant class sampling costs cognitive effort while providing no additional information.

In the Tight Proximity condition, there were physical cues indicating that certain symbols belonged to the same class (i.e., symbols were close together). These grouping cues provided a "boundary" which participants could recognize and utilize easily. Repetitious sampling of same-class items was reduced, thus reducing subjective workload.

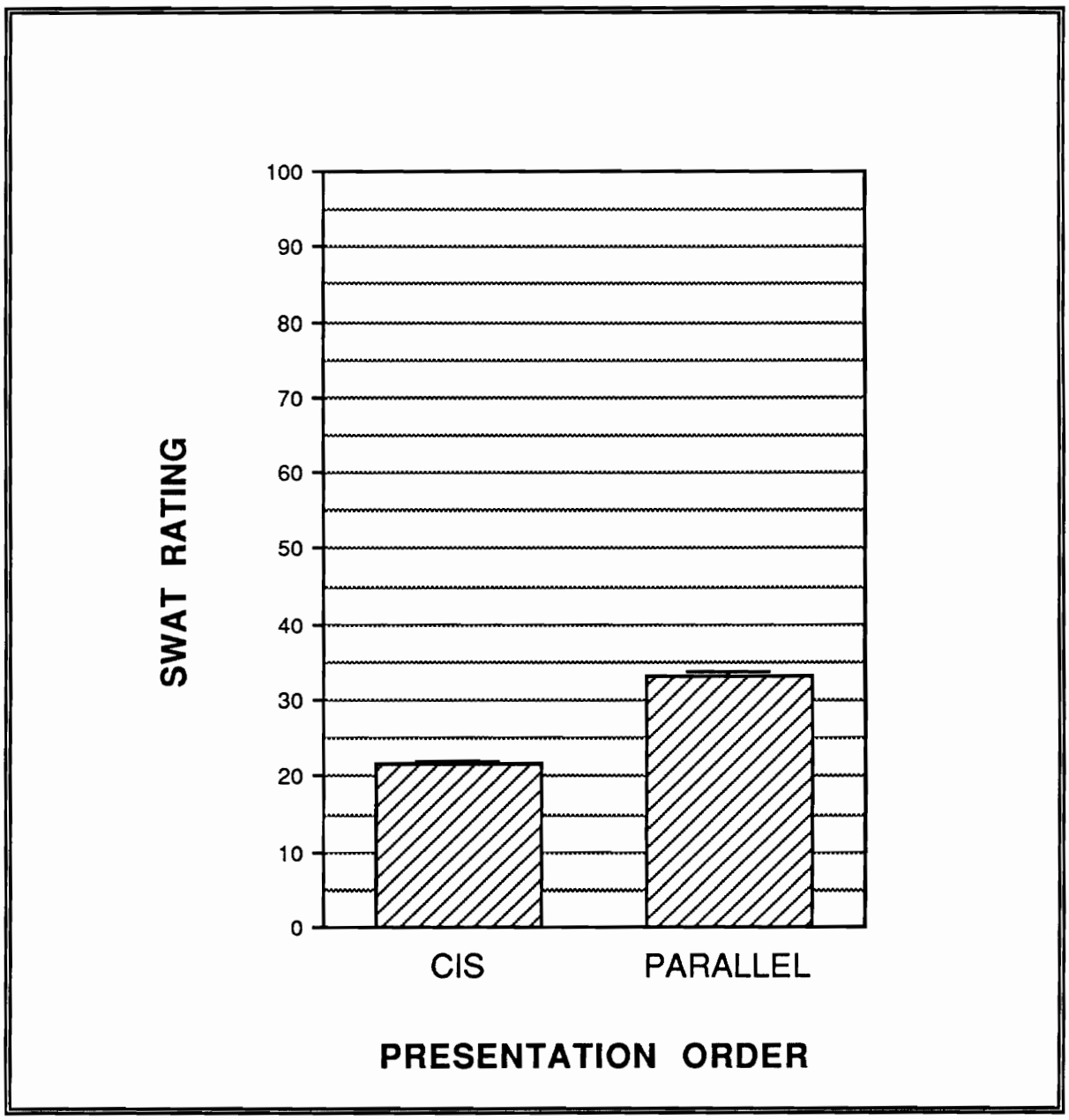


Figure 16. Effect of Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Proximity, and Type, on SWAT ratings in Experiment 1 ($p < 0.0006$). Error bars indicate + 1 $s_{\bar{x}}$.

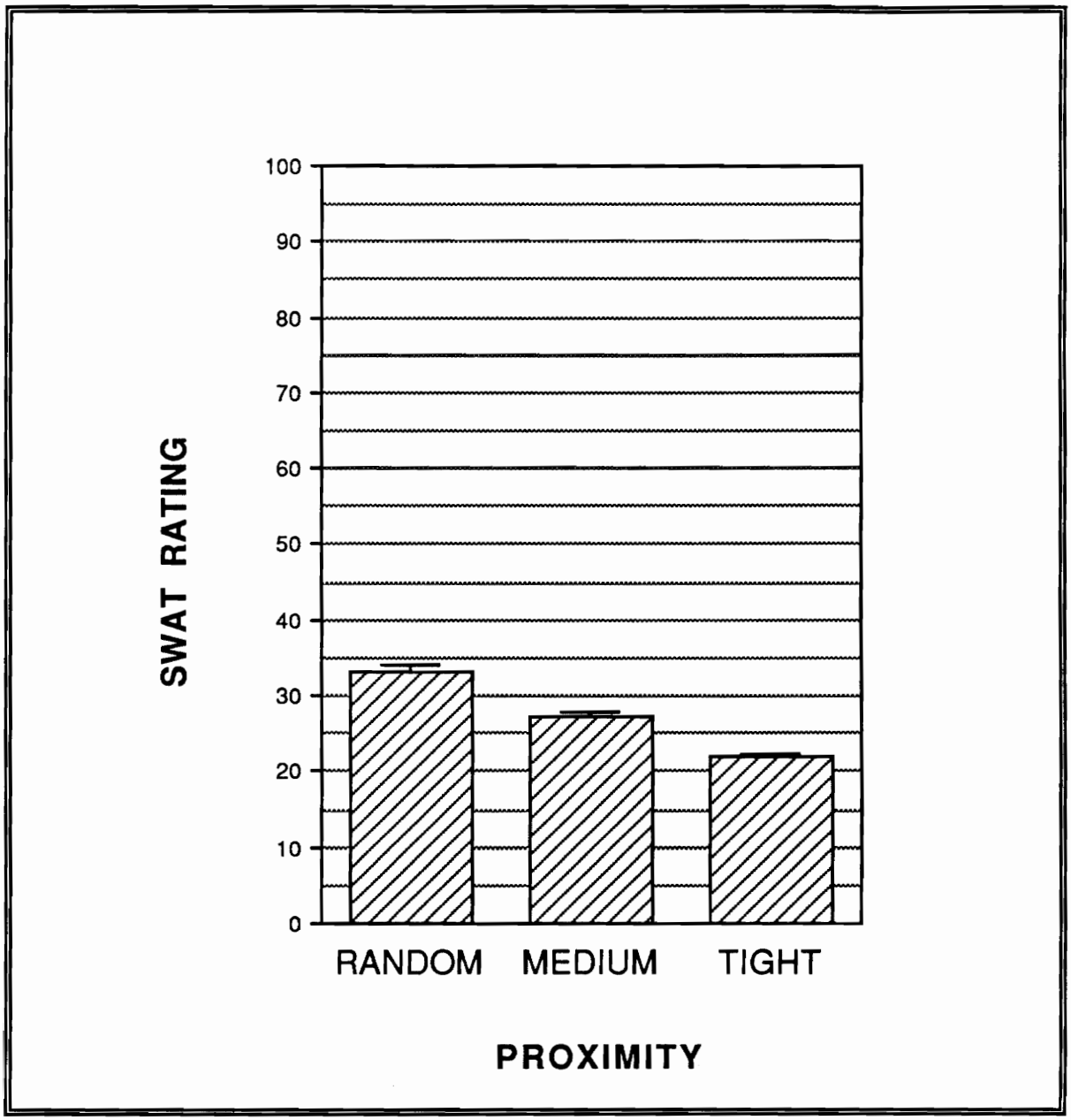


Figure 17. Effect of Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Presentation Order, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars indicate + 1 $s_{\bar{x}}$.

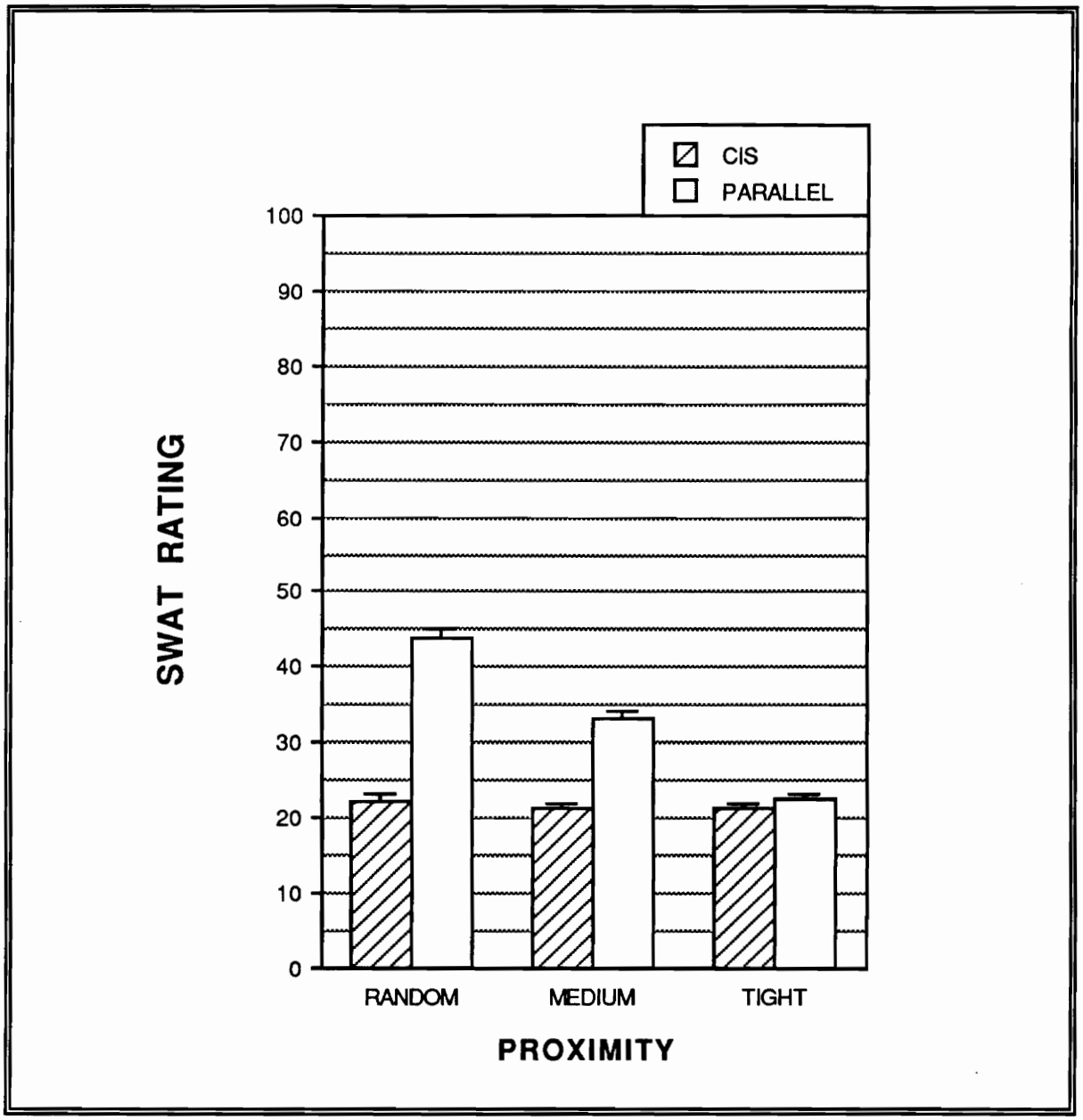


Figure 18. Two-factor interaction of Presentation Order X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars indicate + 1 $s_{\bar{x}}$.

CIS presentations produced lower SWAT ratings than do the Random and Medium-Proximity/Parallel-Order presentations, because same-class symbols were isolated by a temporal boundary. Since participants knew that all symbols presented at any time belonged to the same class, they could have sampled one symbol and known that they had sampled all symbols belonging to that class.

In the CIS presentation, SWAT ratings are similar regardless of Proximity. This is easily explained. In CIS presentations, each class was isolated from the others by a separation in time. Thus, items in various levels of Proximity neither increased nor decreased the amount of separation between classes.

Finally, it is interesting to note that error rates and subjective workload levels are equal for all of the CIS presentations and the Parallel-Order/Tight-Proximity condition. It seems that, for this task, grouping the classes improved performance to an equivalent level regardless of whether this grouping is achieved by a temporal or a spatial method.

Combined effects of complexity and grouping. The two-factor interaction between Number Of Items and Presentation Order is significant ($p < 0.0001$). As shown in Figure 19 (and Table F9), SWAT ratings do not vary with Number Of Items under the Parallel Order condition, but do increase with increasing Number Of Items under the CIS Order. It is suggested that, when the Number Of Items equaled one, the participant saw between 2 and 5 symbols (classes) presented either simultaneously (i.e., Parallel Order) or sequentially (i.e., CIS Order). In this case, each symbol represented a single class and, therefore, there was no display "noise" (i.e., redundant class symbols) which might be reduced by facilitative grouping. Under the CIS Order, increasing the Number Of Items did not influence subjective workload, as the temporal dimension provided effective boundaries between symbol classes. However, under the Parallel presentations, boundaries among classes became less clear as the Number Of Items increased.

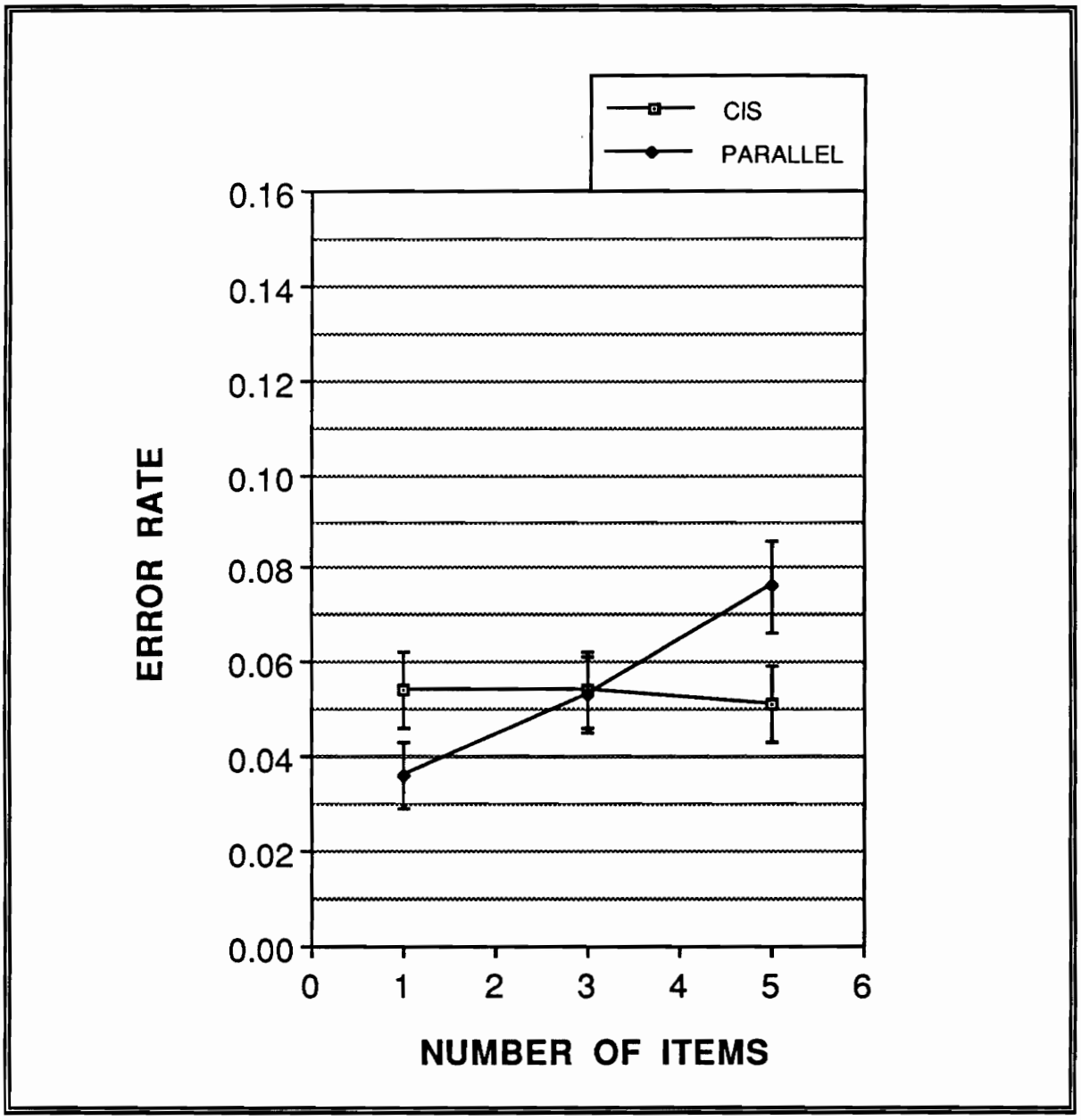


Figure 19. Two-factor interaction of Number Of Items X Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Classes, Proximity, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

The two-factor interaction between Number Of Items and Proximity is significant ($p < 0.0001$). In Figure 20 (and Table F10) it can be seen that Proximity has no effect on SWAT ratings when the Number Of Items equals one. As the Number Of Items increases, workload costs due to low Proximity also increase.

This interaction can be explained as follows: first, when the Number Of Items equals one, Proximity has no effect. The reader is reminded that Spatial Proximity describes the spatial relationship of symbols belonging to the same class. When there is only one item per class, displays are equivalent regardless of the spatial Proximity. It is reasonable to expect that display strategies which group same-class items will be effective only when there is more than one item per class. Second, when Proximity was tight, the spatial grouping provided easily discernable boundaries between symbol classes, so participants could extract the class name regardless of the Number Of Items per class. However, as Proximity decreased participants had either more ambiguous or non-existent spatial boundaries to help chunk symbols belonging to the same class. In this case, increases in the Number Of Items per class lead to increases in task difficulty.

The three-factor interaction among Number Of Items, Presentation Order, and Proximity is significant ($p < 0.0001$) (Figure 21). In general, this three-way interaction is fully consistent with its component lower-order interactions. For example, a series of simple-effects F-tests (Table F11) indicates that the two-factor interaction of Presentation Order and Proximity is not significant when Number Of Items equals one ($p > 0.05$), but is significant when the Number Of Items equals three or five ($p < 0.01$). Newman-Keuls tests (Tables F12 through F14) indicate that when the Number Of Items equals one, SWAT ratings do not vary significantly with Presentation Order or Proximity. This is consistent with the Number of Items X Presentation Order and Number Of Items X Proximity

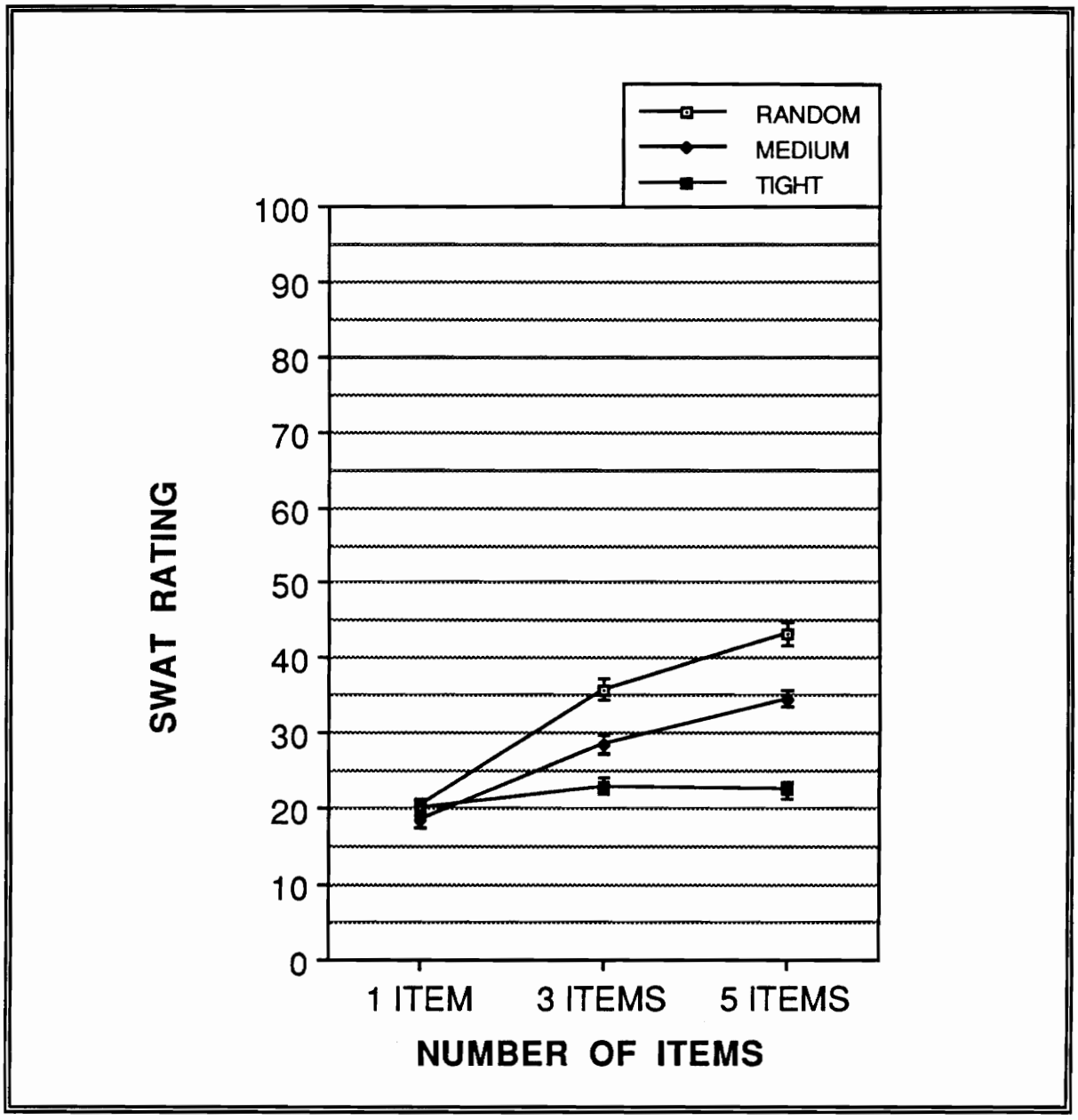


Figure 20. Two-factor interaction of Number Of Items X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

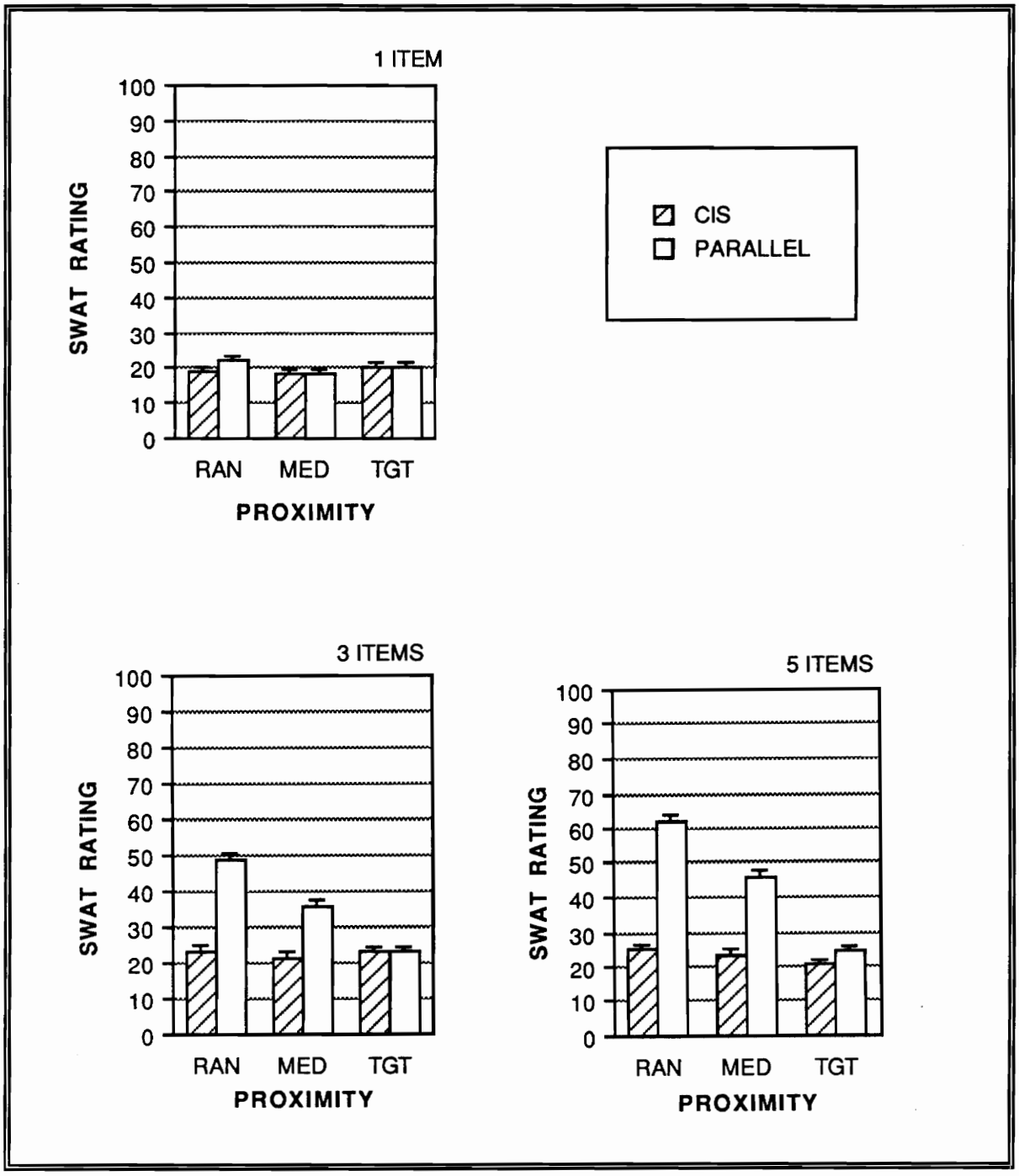


Figure 21. Three-factor interaction of Number Of Items X Presentation Order X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars indicate + 1 $s_{\bar{x}}$.

interactions. However, when the Number Of Items equals either three or five, results are consistent with the Presentation Order X Proximity interaction. Specifically, when (1) Presentation Order is CIS, Proximity does not affect SWAT ratings (2) when Presentation Order is Parallel, SWAT ratings increase as Proximity decreases and (3) the Parallel-Order/Tight-Proximity condition provides SWAT ratings similar to CIS presentations. It is noted that SWAT differences associated with Presentation Order and Proximity increase as the Number Of Items increases from 3 to 5.

The three-factor interaction among Number Of Classes, Number Of Items, and Proximity is significant ($p < 0.0574$). Figure 22 illustrates that, when the Number of Items equals 3 or 5, Proximity effects became less marked as the Number of Classes increases. Secondly, as the Number Of Classes increases, workload associated with the 1-item level becomes more similar in overall magnitude to that associated with the 3- and 5-item levels. It seems that subjective workload contributed by increasing numbers of classes overshadows the workload contributed by the Number Of Items and the Proximity levels.

Finally, it is reasonable that the Number Of Items interacts with display grouping strategies, while the Number Of Classes does not. As suggested earlier, the Number Of Items impacts only display confusion (i.e., noise-to-signal ratio) while the number of Classes impacts MSET size, a factor which is not enhanced by facilitative grouping formats.

Effects due to Type. Type has a significant effect on SWAT ratings ($p < 0.0194$) (Figure 23). Average ratings are 28.53 for trials requiring negative responses and 26.29 for trials requiring positive responses. The two-factor interaction of Number Of Items X type is significant ($p < 0.0369$). As illustrated in Figure 24 (and Table F 15), the disparity between negative and positive responses increases with increasing Number Of Items.

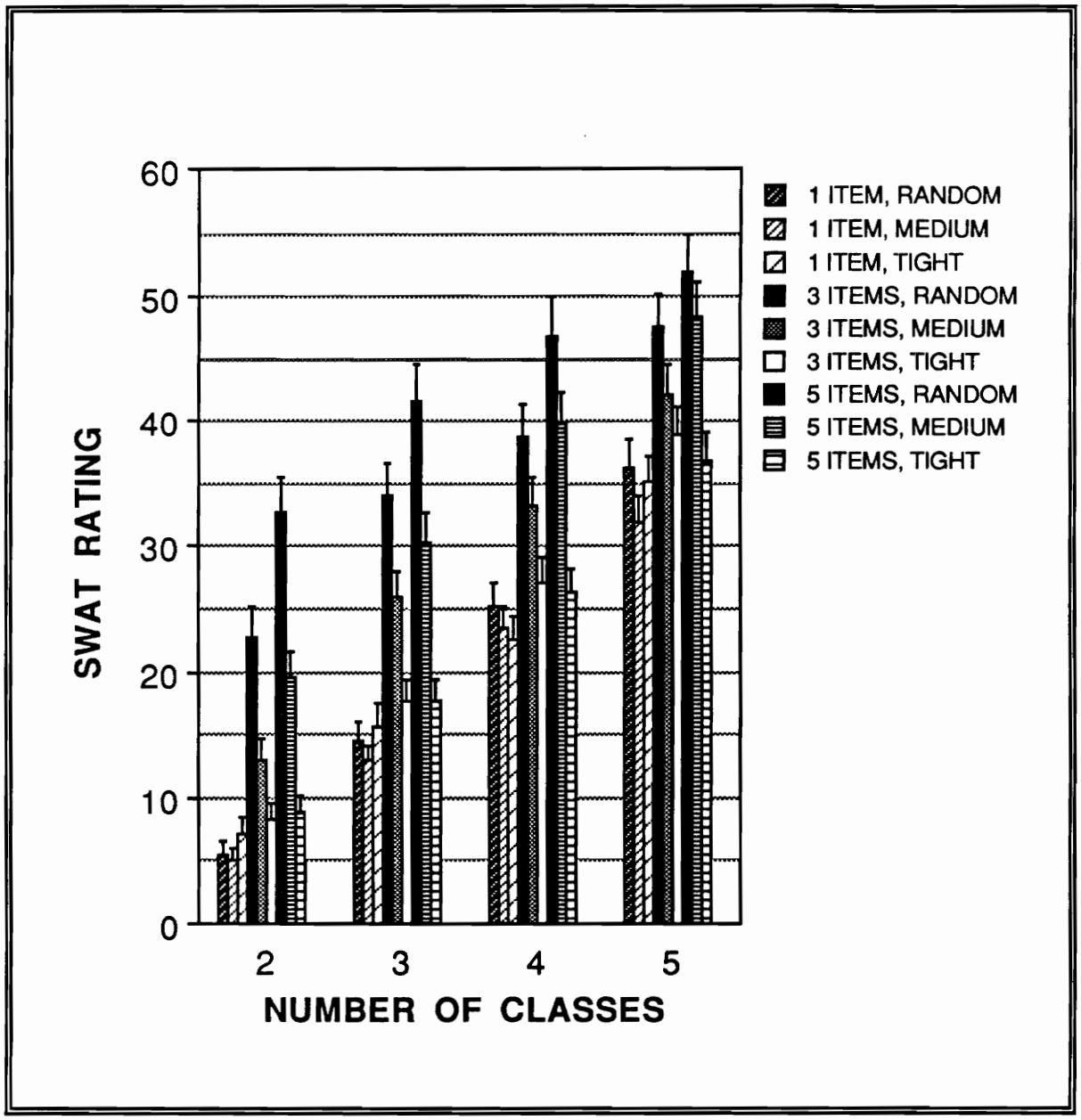


Figure 22. Three-factor interaction of Number Of Classes X Number Of Items X Proximity, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 1 ($p < 0.0574$). Error bars indicate + 1 $s_{\bar{x}}$.

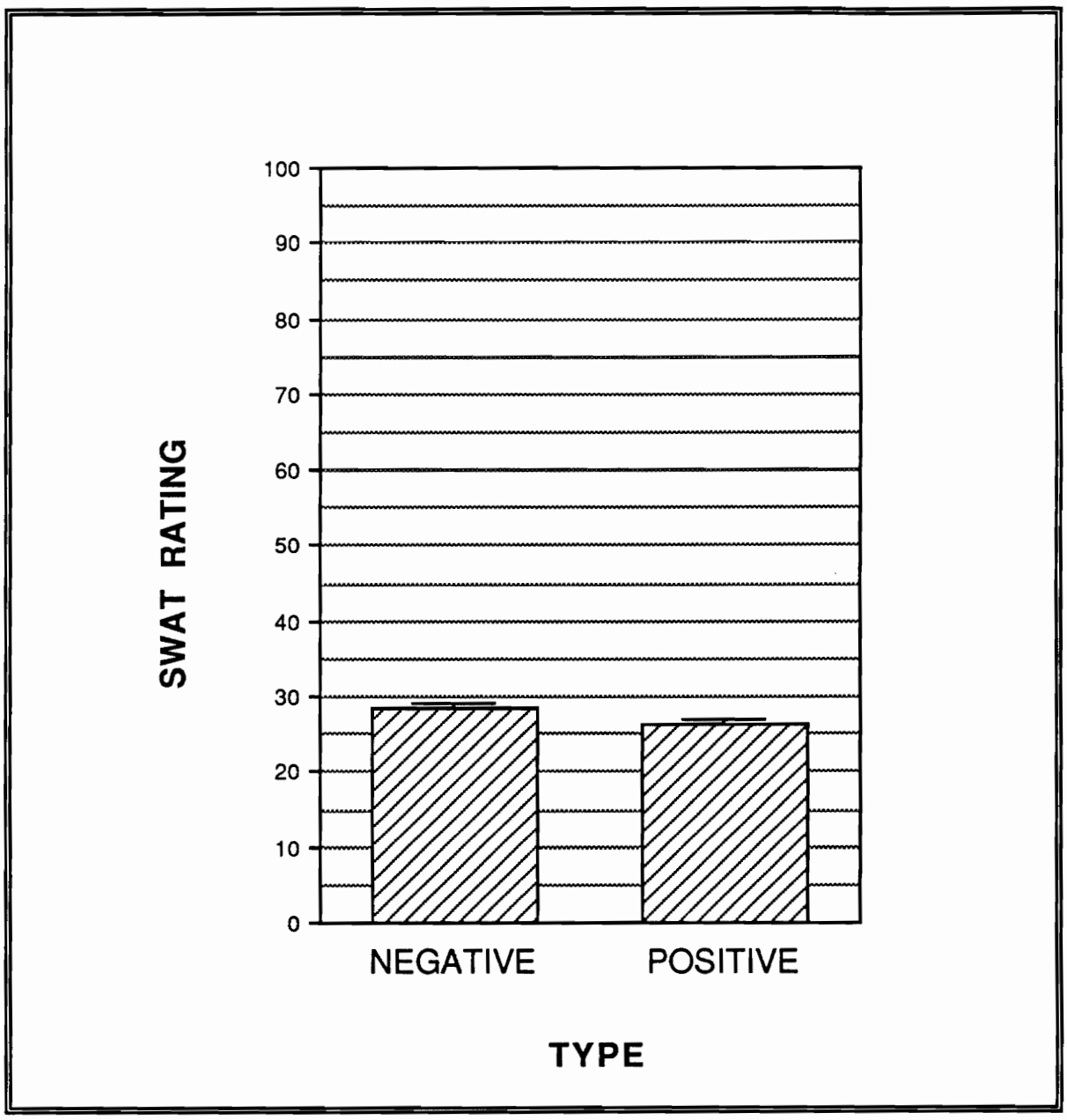


Figure 23. Effect of Type, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Presentation Order, and Proximity, on SWAT ratings in Experiment 1 ($p < 0.0194$). Error bars indicate + 1 $s_{\bar{x}}$.

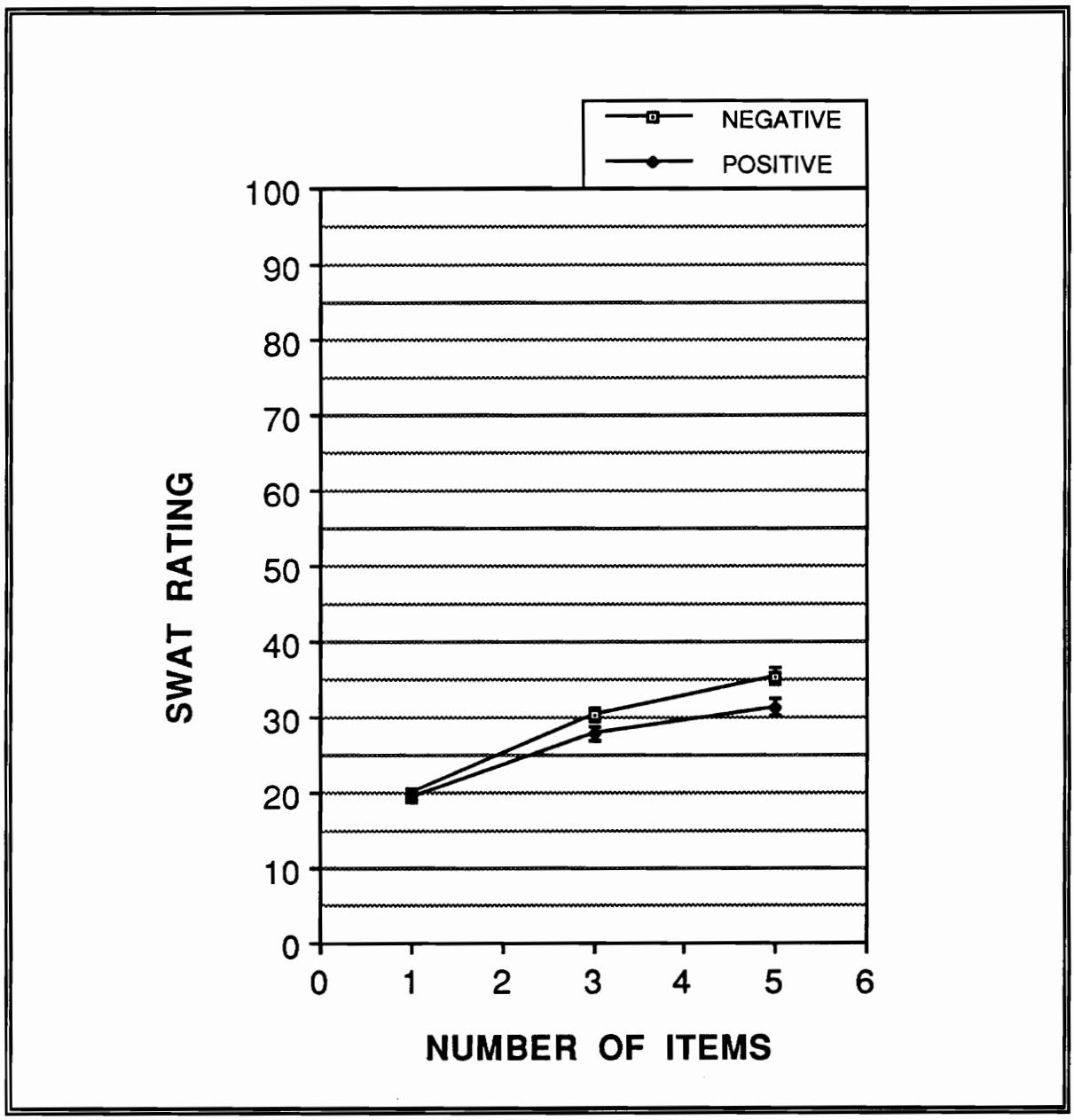


Figure 24. Two-factor interaction of Number Of Items X Type, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, and Proximity, on SWAT ratings in Experiment 1 ($p < 0.0369$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

It is suggested that in a positive response trial, a participant could match the probe to an MSET item. He could then terminate his search and respond "yes". When failing to make a positive identification, a participant's search would have been exhaustive of all known MSET items. Second, upon failing to make a positive match, some participants might have been less sure of their responses and judged the trial as more difficult.

Finally, the three-factor interaction among Number Of Classes, Presentation Order, and Type is significant ($p < 0.0470$). As shown in Figure 25, this three-factor interaction is consistent with the component two-factor interactions. In the CIS Order, the effects of Type become more marked as the Number Of Classes increased. In the Parallel Order, the converse effect is observed; that is, differences become less marked as the Number Of Classes increases.

Error Rate Data

Of the 4320 trials, participants made 234 errors, accounting for 5.4 percent of the trials. In comparison, Sternberg's (1966, Experiment 1) participants made errors on 1.4 percent of the trials. However data for 3 of Sternberg's 11 participants were not analized as they exceeded an error criterion. (This error criterion was not reported.)

Effects due to display complexity. As shown in Figures 26 and 27 (and Tables F16 and F17), error rate increases monotonically with increasing Number Of Classes ($p < 0.0001$) and Number Of Items ($p < 0.0615$), respectively. Comparing these figures, the effect of Number Of Classes appears pronounced compared to that due to the Number Of Items.

The two-factor interaction between the Number Of Classes and the Number Of Items is significant ($p < 0.048$). As shown in Figure 28, error rate increases as a function of the Number Of Classes and Number Of Items, yet the specific trend is difficult to identify. As an auxiliary analysis, the effect of the total number of displayed symbols (i.e., Number Of

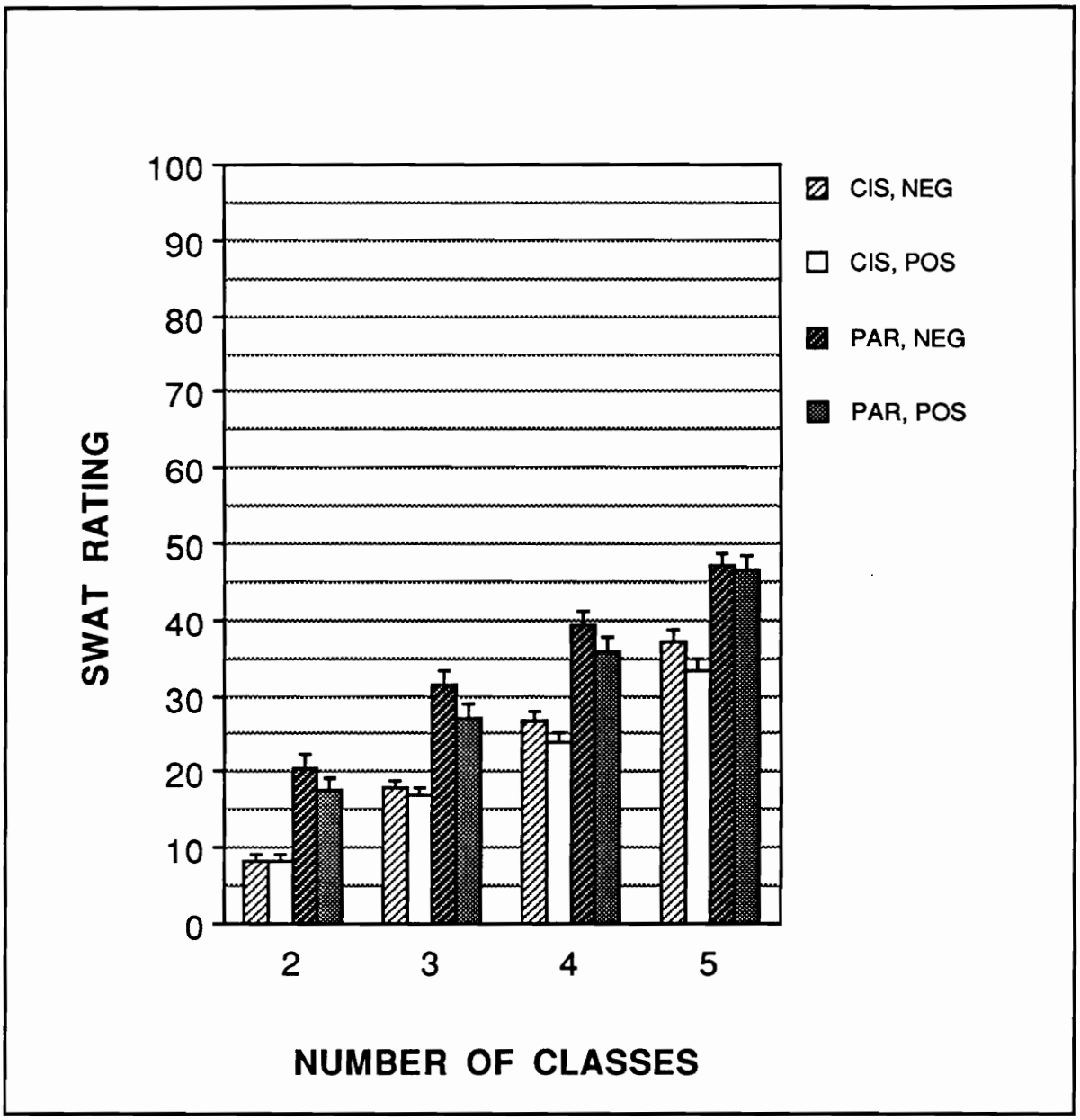


Figure 25. Three-factor interaction of Number Of Classes X Presentation Order X Type, collapsed over subjects (N = 10), Blocks, Number Of Items, and Proximity, on SWAT ratings in Experiment 1 ($p < 0.0369$). Error bars indicate + 1 $s_{\bar{x}}$.

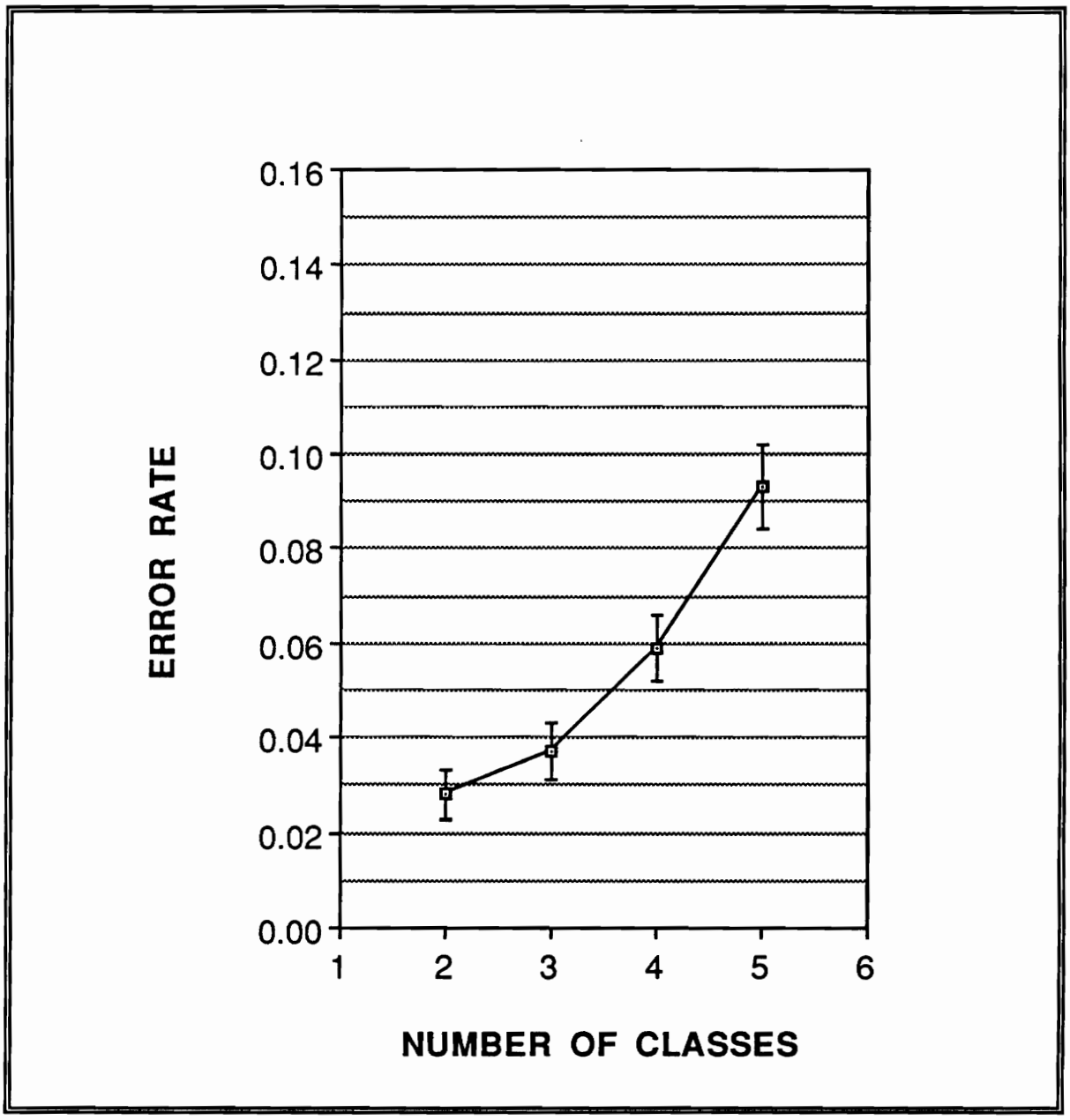


Figure 26. Effect of Number Of Classes, collapsed over subjects (N = 10), Blocks, Number Of Items, Presentation Order, Proximity, and Type, on error rates in Experiment 1 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

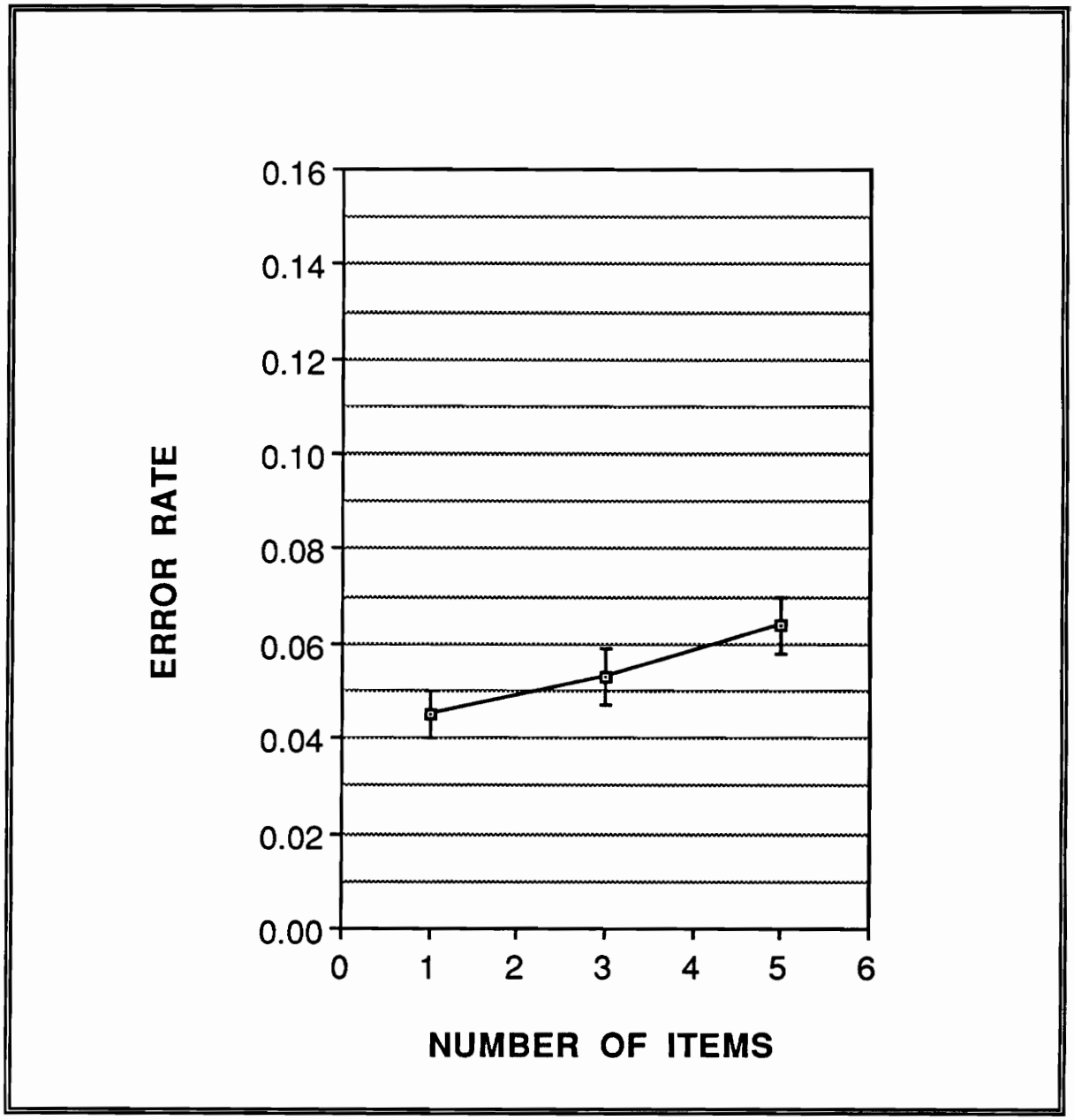


Figure 27. Effect of Number Of Items, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, Proximity, and Type, on error rates in Experiment 1 ($p < 0.0615$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

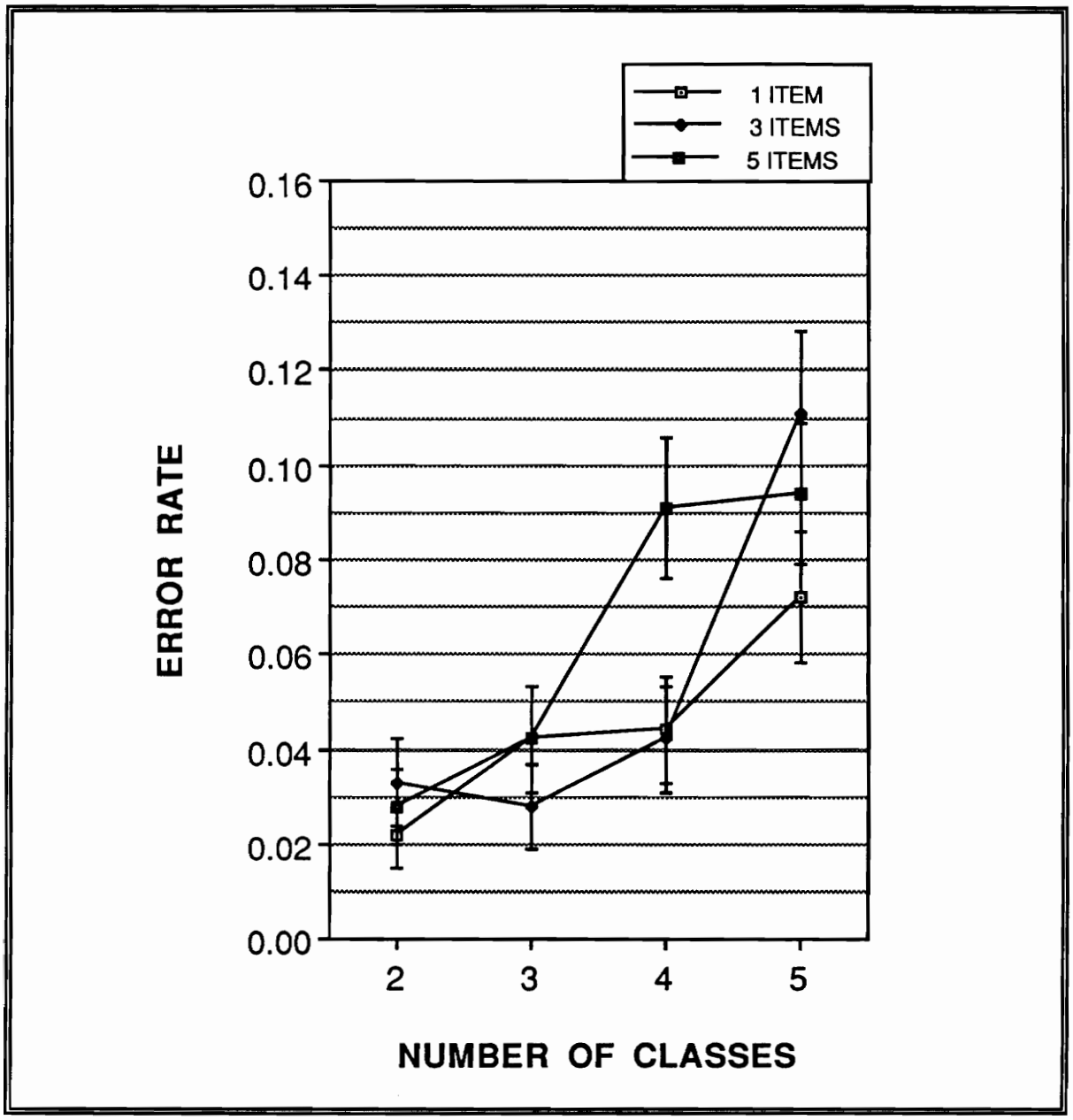


Figure 28. Effect of Number Of Classes X Number Of Items, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on error rates in Experiment 1 ($p < 0.0478$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

Classes x Number Of Items) displayed during any one trial is shown (Figure 29). As observed in the SWAT analysis, the effect of total symbols on error rate shows a bimodal trend. Although a Newman-Keuls test (Table F17) indicates few statistically significant differences, the correspondence of data trends across dependent measures points up the consistency of this effect.

Effects due to display grouping. As shown in Figure 30, error rate decreases monotonically as proximity increases ($p < 0.0166$). A Newman-Keuls test (Table F18) indicates that some, but not all Proximity levels differ significantly.

The two-factor interaction between Presentation Order and Proximity, shown in Figure 31, is significant ($p < 0.0035$). A Newman-Keuls test (Table F19) indicates that more errors occur in the Parallel-Order/Random-Proximity combination than in any other condition. Also, error rate differences among all other conditions do not differ significantly. Since the CIS presentations are associated with fewest errors across all Proximity conditions, this data trend closely corresponds to the analogous SWAT findings (Figure 18). Again, it is suggested that CIS provides observers with easily identified class boundaries and helps reduce redundant sampling of same-class symbols.

Combined effects of complexity and grouping. As shown in Figure 32, the two-factor interaction between the Number Of Classes and Presentation Order is significant ($p < 0.0067$). A Newman-Keuls test (Table F20) indicates that the four-class/Parallel, five-class/Parallel, and five-class/CIS combinations do not differ significantly, but do differ from all other combinations. The remaining combinations are not statistically different from one another ($p > 0.05$).

The two-factor interaction between Number Of Items and Presentation Order is significant ($p < 0.0667$). As shown in Figure 33, error rate is not affected by the Number Of Items under the CIS Order; however, error rates increase with increasing Number Of

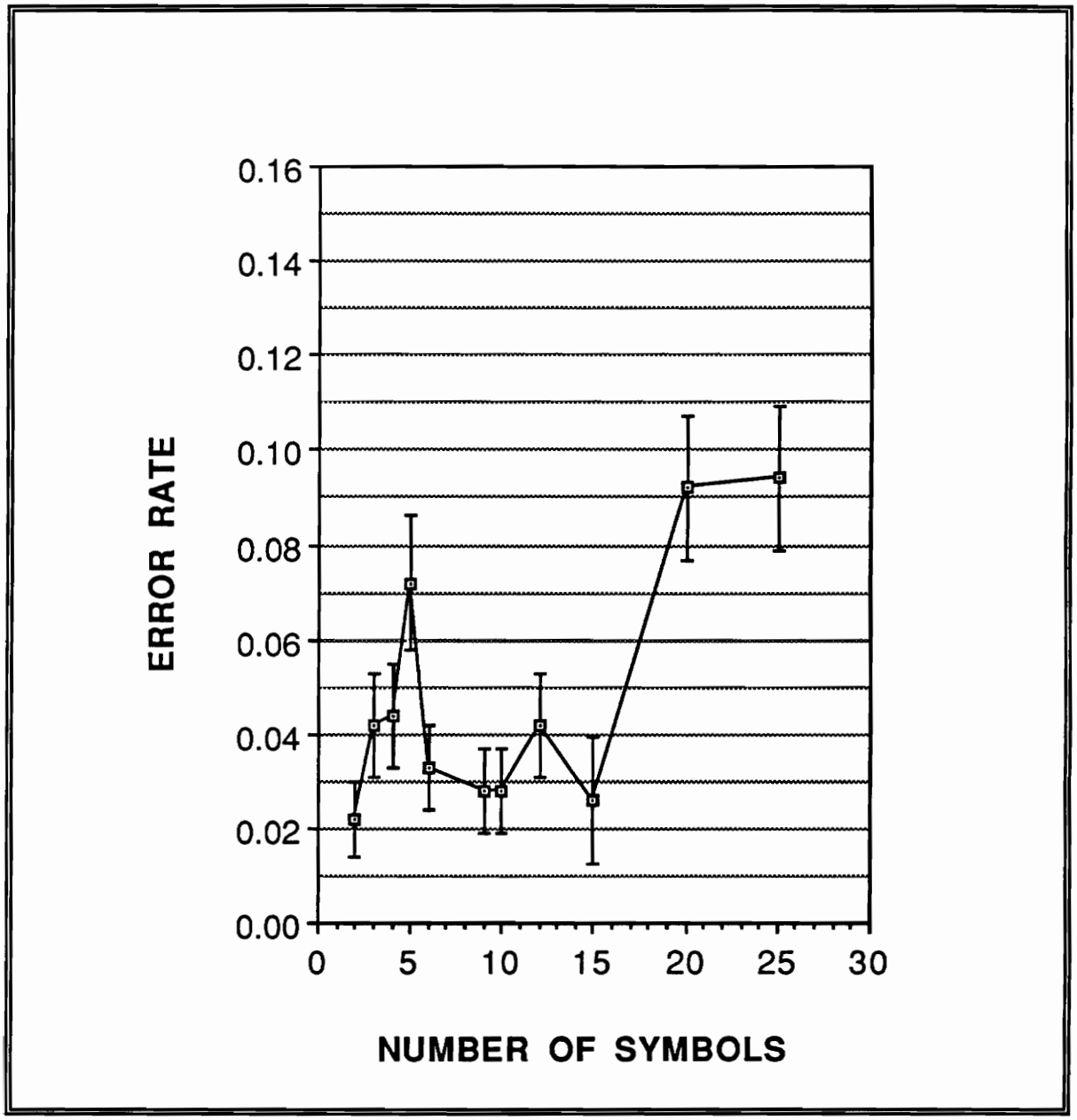


Figure 29. Effect of number of symbols, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on error rates in Experiment 1. Error bars indicate $\pm$ 1 $s_{\bar{x}}$. Note: There were two cases in which the number of symbols equalled 15. Average means and $s_{\bar{x}}$ s are shown.

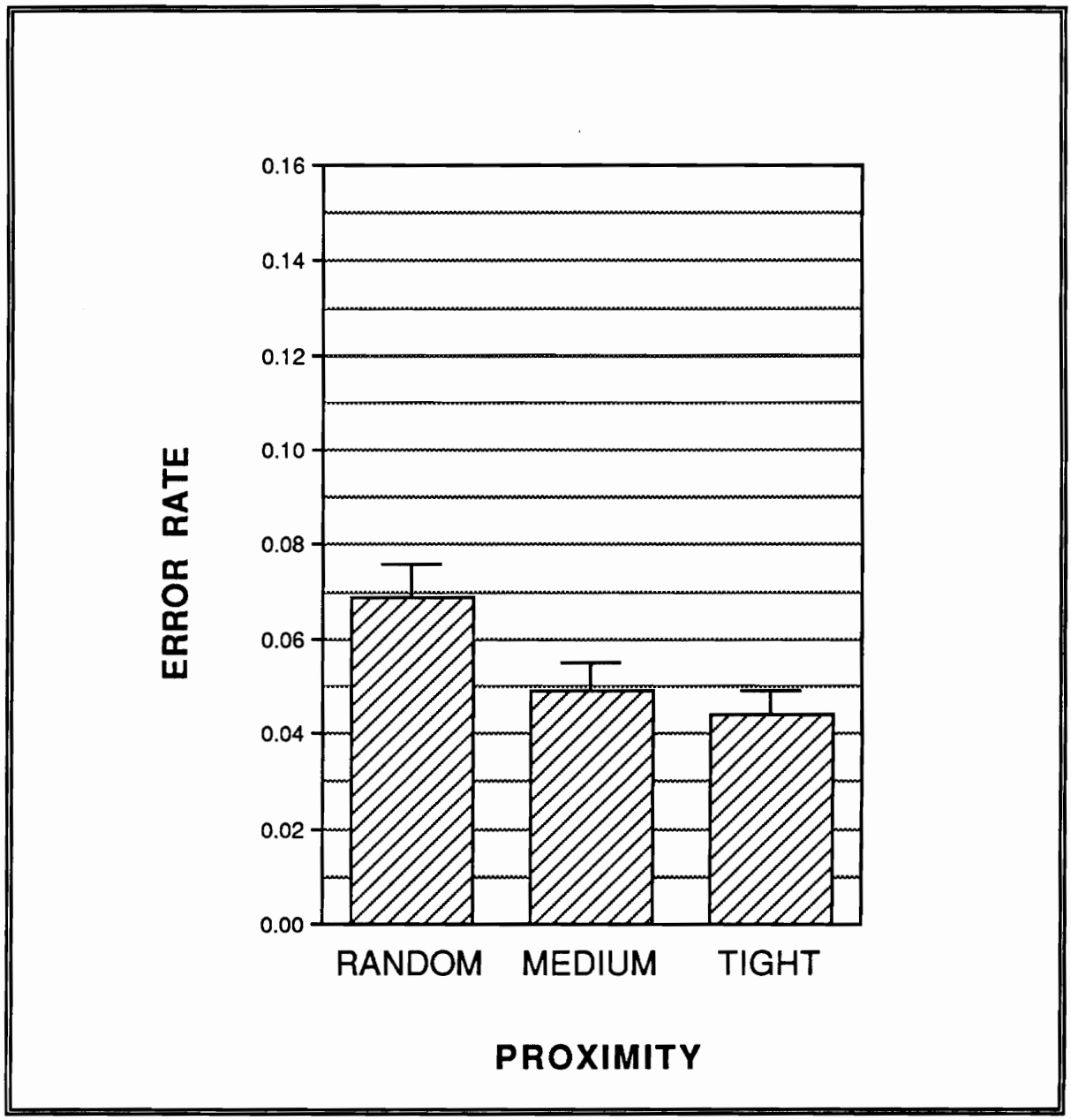


Figure 30. Effect of Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Presentation Order, and Type, on error rates in Experiment 1 ($p < 0.0166$). Error bars indicate + 1 $s_{\bar{x}}$.

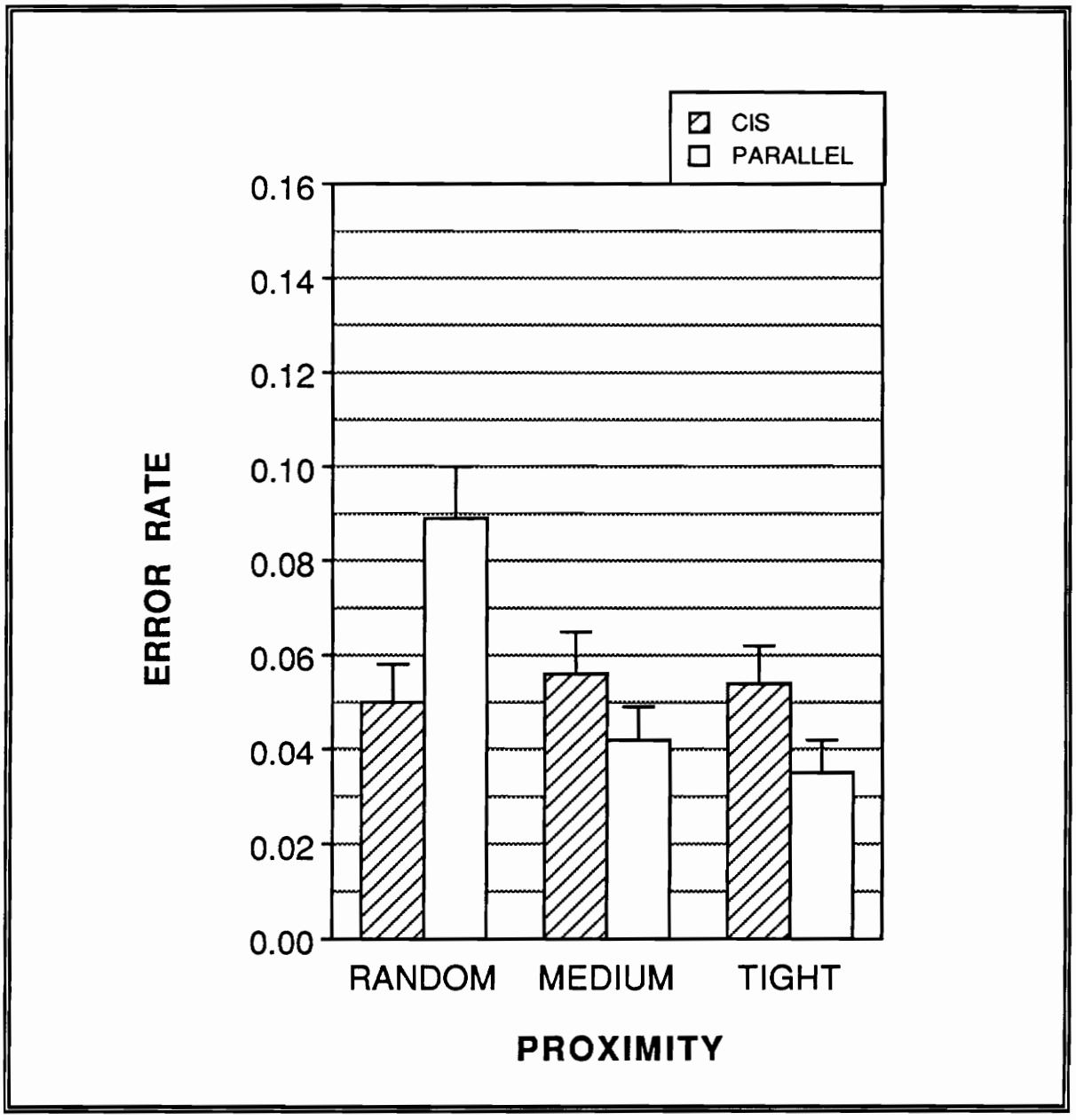


Figure 31. Two-factor interaction of Presentation Order X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, and Type, on error rates in Experiment 1 ($p < 0.0035$). Error bars indicate + 1 $s_{\bar{x}}$.

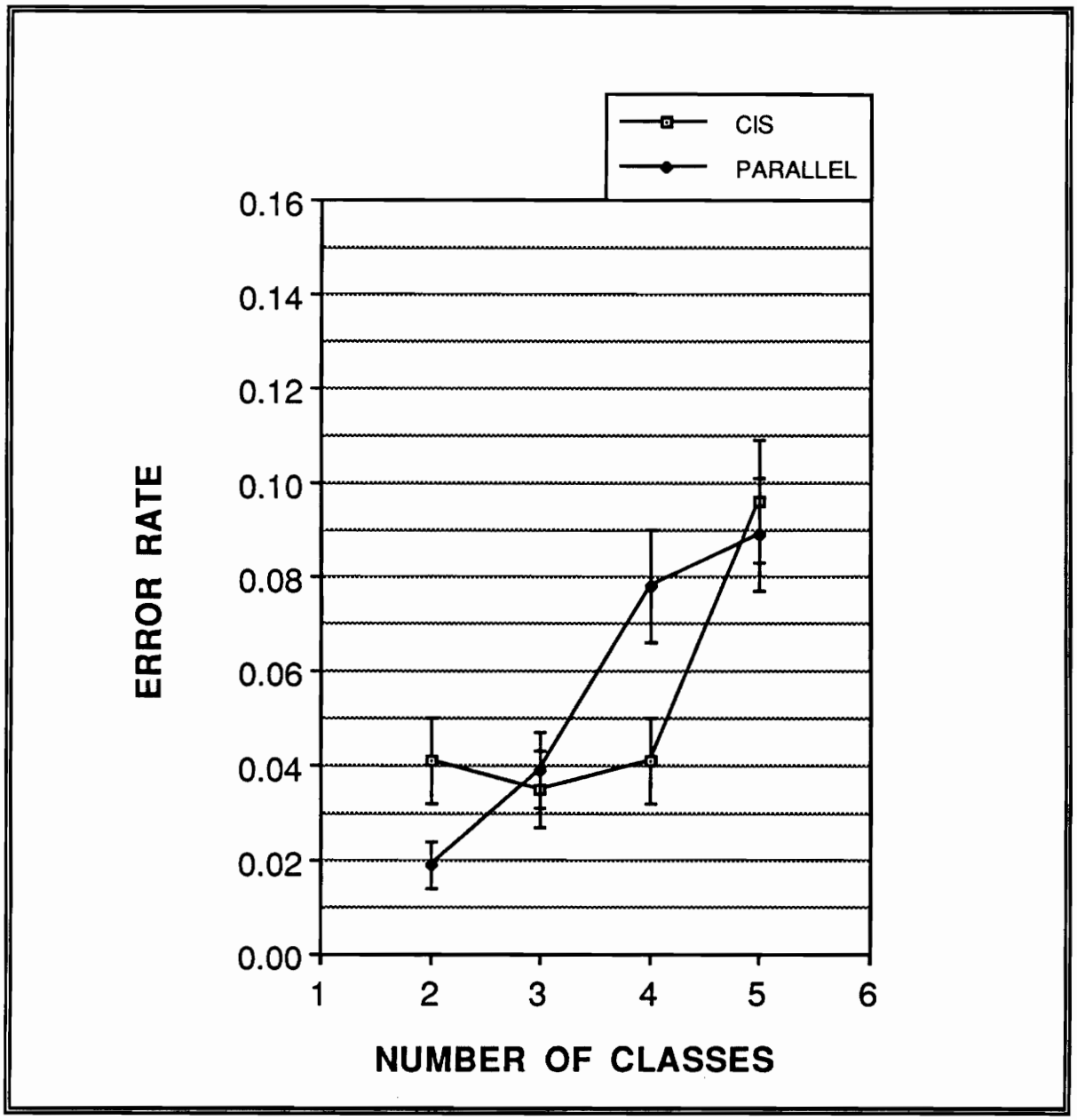


Figure 32. Two-factor interaction of Number Of Classes X Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Items, Proximity, and Type, on error rates in Experiment 1 ($p < 0.0067$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

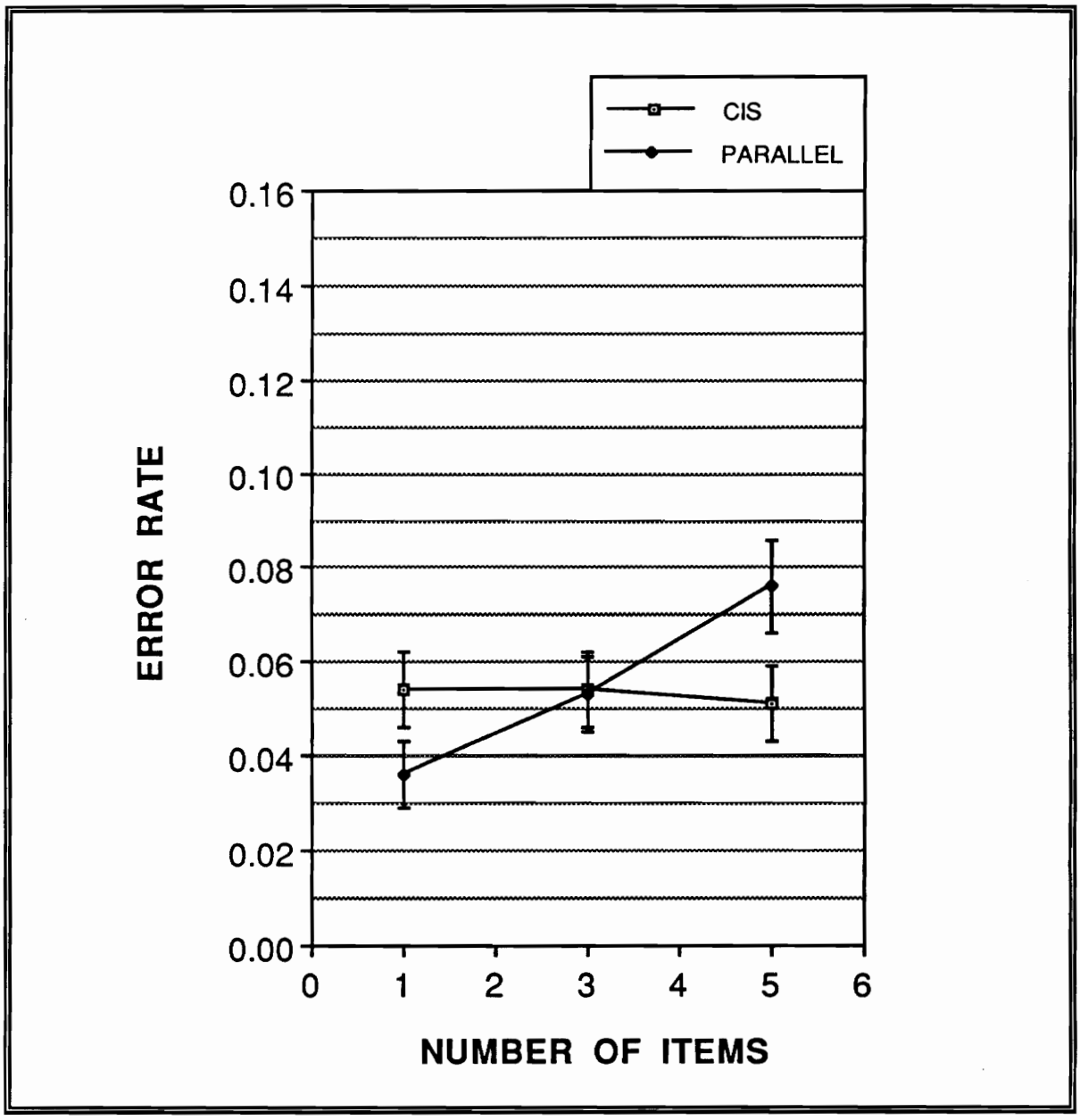


Figure 33. Two-factor interaction of Number Of Items X Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Classes, Proximity, and Type, on error rates in Experiment 1 ($p < 0.0666$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

Items under the Parallel Order. It is important to note from this effect that error rate under the CIS presentation mode is largely unaffected by increasing Number Of Items.

Second, error rate is higher for CIS Order when the Number Of Items equals five, while error rate is higher for Parallel Order when the Number Of Items equals one. It is suggested that CIS Order provides easily discernable boundaries between classes. These boundaries produce increasing benefits as the noise-to-symbol ratio increases. When the Number Of Items was one, CIS classes (i.e. display screens) were presented then blanked every 0.8 s. It is possible that, on some trials, participants had insufficient time to encode the symbols and assign class labels.

Finally, the three-factor interaction, shown in Figure 34, among Number Of Classes, Presentation Order, and Proximity is significant ($p < 0.063$). In general, the data trends observed in this interaction effect do not modify the findings based on its component (i,.e., lower order) effects. Clearly, the important finding that error rate remains low under CIS Presentation Order is evident in these data. Also, it can be seen that error rate increases in Parallel Presentation Order with increasing Number Of Classes or decreasing Spatial Proximity. Given the nature of these trends, the large $s_{\bar{x}}$.and degrees of freedom in these data, the presence of this statistically significant effect is believed to be unimportant.

These interactions were examined to determine whether display grouping formats provide greater error reduction as display complexity increases. However, there is no discernable, consistent pattern across these interactions. Therefore, it is prudent to conclude that, given this task, display grouping formats do not alter error rates across display complexity levels.

Effects due to Type. Two-factor interactions between the Number Of Items and Type ($p < 0.0261$) as well as between the Number Of Classes and Type ($p < 0.0735$) are

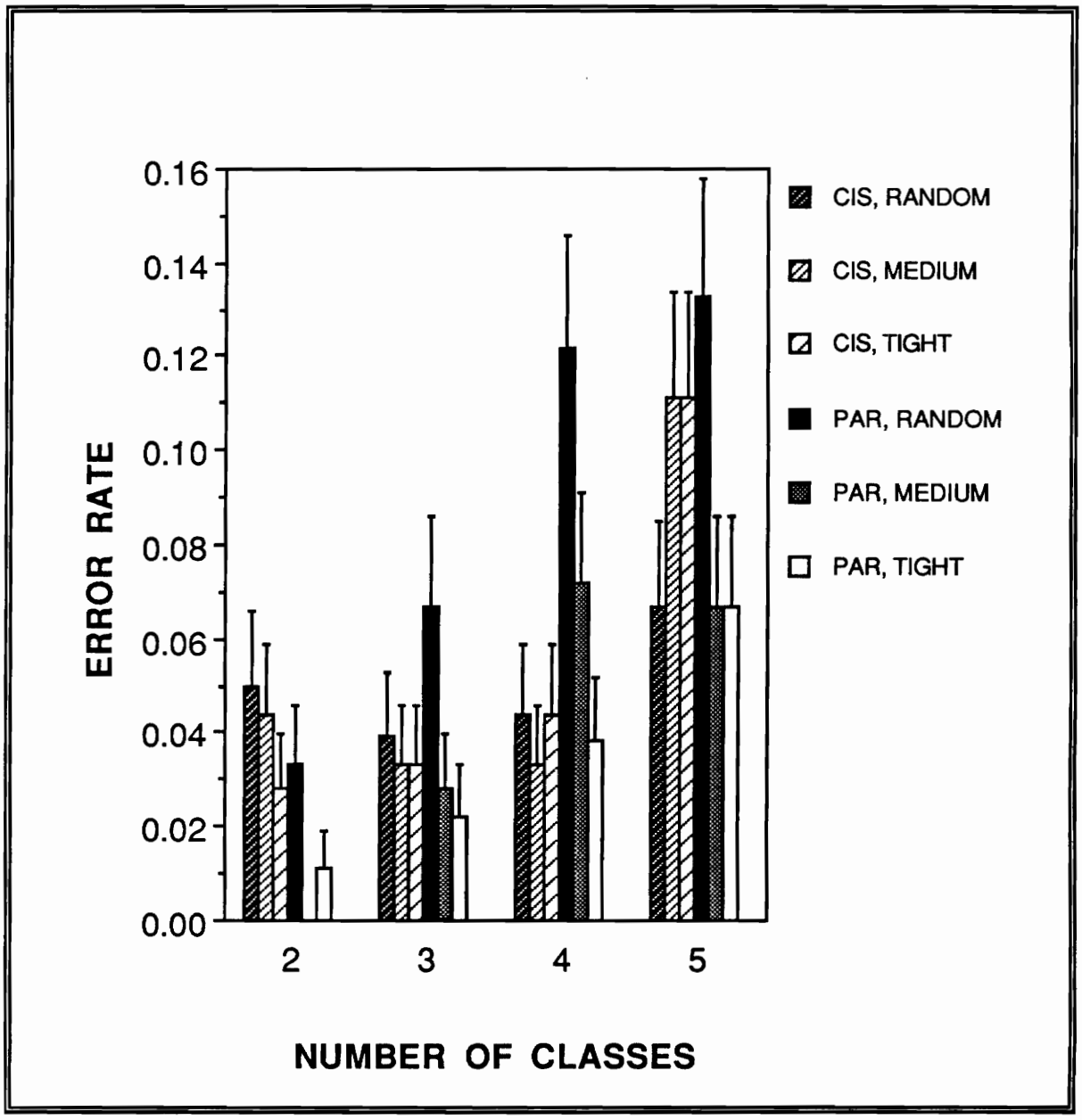


Figure 34. Three-factor interaction of Number of Classes X Presentation Order X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Items, and Type, on error rates in Experiment 1 ($p < 0.0666$). Error bars indicate + 1 $s_{\bar{x}}$.

significant. Considering only the Number Of Items X Type interaction shown in Figure 35 (and Table F21), error rate increases with increasing Number Of Items under the Negative Probe Type condition, but remains relatively low under the Positive Type condition. A similar trend of Probe Type is observed across Number Of Classes levels (not shown).

One explanation of these interactions can be derived from post-experiment debriefings. Upon positively identifying an MSET item, participants respond "yes", but when failing to make a positive identification, participants made educated guesses. Participants indicated that their guesses were based upon information such as whether they thought most of the displayed shapes had been rounded or angular. Thus, it seems reasonable that participants lost confidence in their responses, particularly under the highest Number Of Items condition.

The three-factor interaction among Number Of Classes, Presentation Order, and Type shown in Figure 36 is significant ($p < 0.0527$). However, the observed data trend does not contradict nor modify the findings based on the lower-order interactions presented above.

Response Time

The average time across all conditions examined was 844.8 ms.

Comments on the response time metric. The emphasis of the Experiment 1 task was subtly different from that of traditional Sternberg tasks. In a traditional Sternberg task, interest rests primarily upon the time required by observers to complete the stimulus decoding process. In Experiment 1, however, treatment variables impacted participants' ability to encode information. That is, a participant had to sift through information to create an MSET list of class labels, and the display grouping levels impacted this process directly. Thus, although response time measures the time required to decode (i.e., scan) the MSET

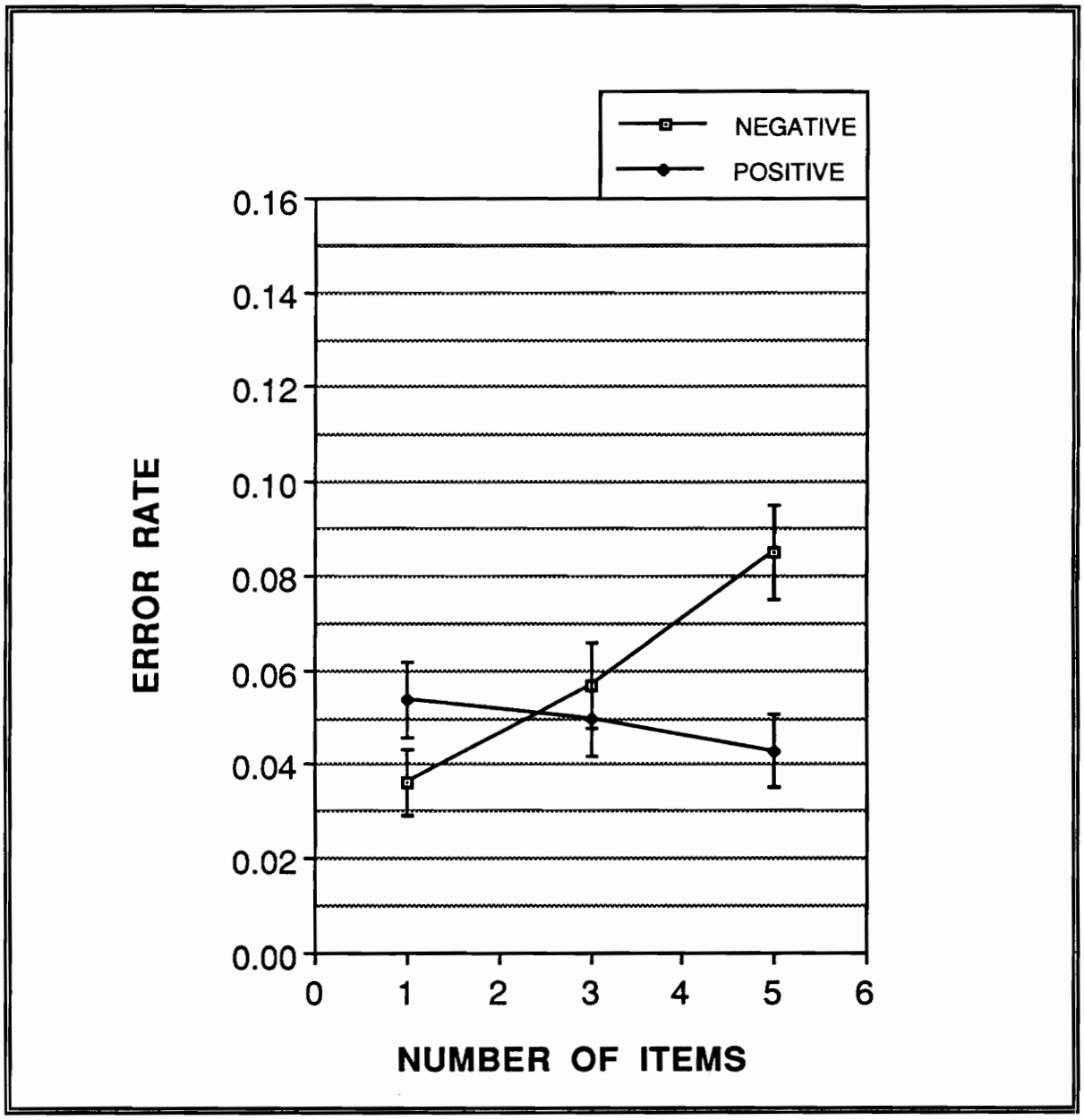


Figure 35. Two-factor interaction of Number Of Items X Type, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, and Proximity, on error rates in Experiment 1 ($p < 0.0261$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

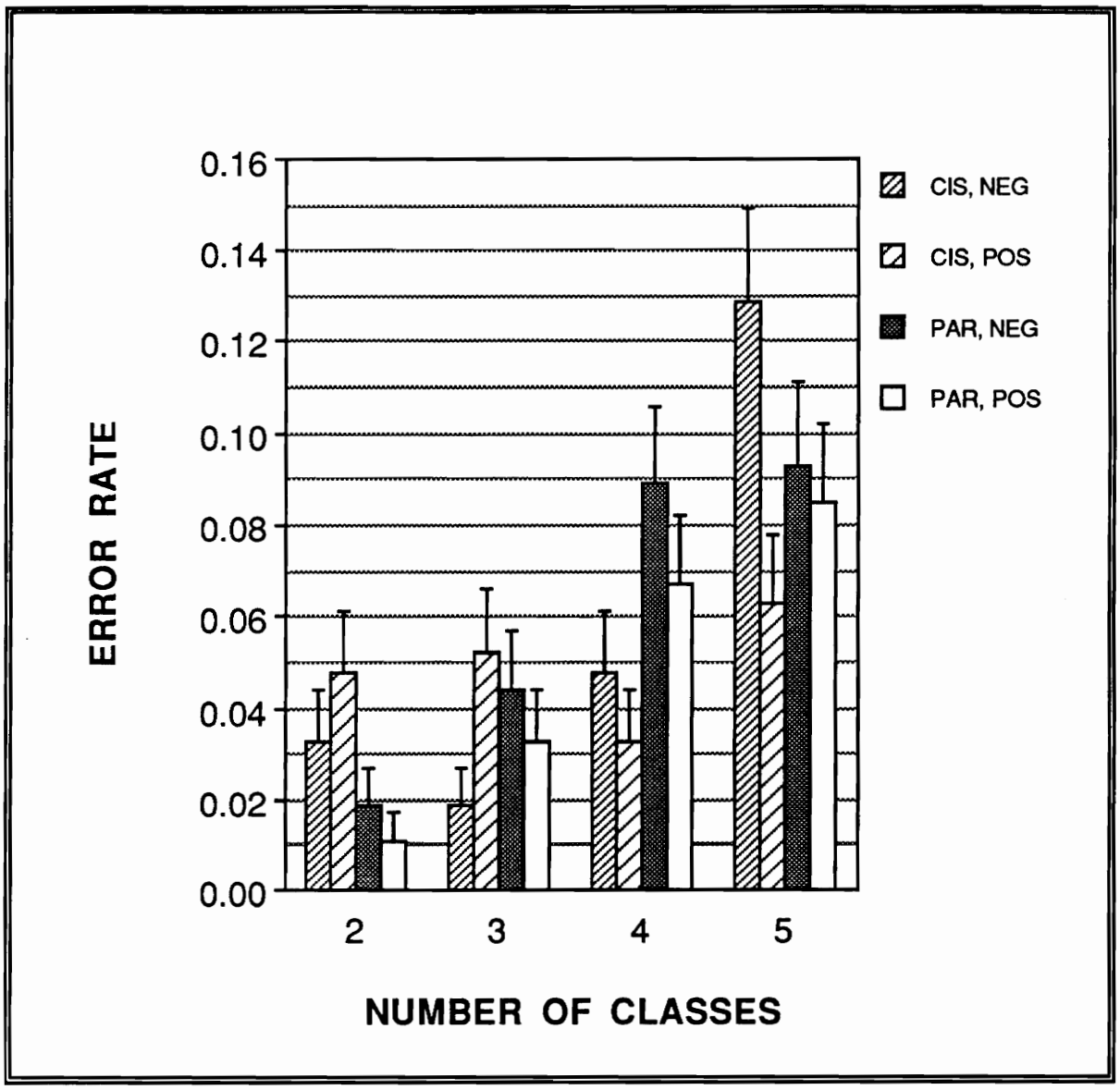


Figure 36. Three-factor interaction of Number Of Classes X Presentation Order X Type, collapsed over subjects (N = 10), Blocks, Number Of Items, and Proximity, on error rates in Experiment 1 ($p < 0.0527$). Error bars indicate + 1 $s_{\bar{x}}$.

list, it is relatively insensitive to MSET encoding mechanisms. The data support this contention.

Effects due to display complexity. As illustrated in Figure 37, response time increases with increasing Number Of Classes ($p < 0.0001$). A Newman-Keuls test (Table F22) indicates that all means are significantly different from one another ($p < 0.05$). The finding that Number Of Classes, but not Number Of Items, affects response time is consistent with previous conclusions. Specifically, Number Of Classes is believed to impact response times since participants scanned each class label (i.e., MSET item) to determine a solution. Conversely, the Number Of Items affects the difficulty of forming an MSET chunk in a manner analogous to signal-to-noise processes. Since MSET chunks are formed prior to probe presentation, the affects of Number Of Items are not well indexed by response time measures.

Effects due to display grouping. Response times are not impacted by levels of display grouping. Again, display grouping primarily affects the extraction of class labels, which was completed prior to probe presentation. Display grouping does not affect the rate at which class labels can be scanned.

Combined effects of complexity and grouping. The Number Of Classes X Proximity interaction is significant ($p < 0.0232$). As illustrated in Figure 38, although reaction time clearly increases with increasing Number Of Classes, the trends attributable to proximity are ambiguous. Since the impact of Proximity is small compared to that of Number Of Classes, the presence of this two-factor interaction does not seriously modify the Number Of Classes main effect findings.

The Number Of Classes X Number Of Items X Presentation Order interaction is significant ($p < 0.0452$) (Figure 39). Except for one condition (i.e., 5-Classes/5-

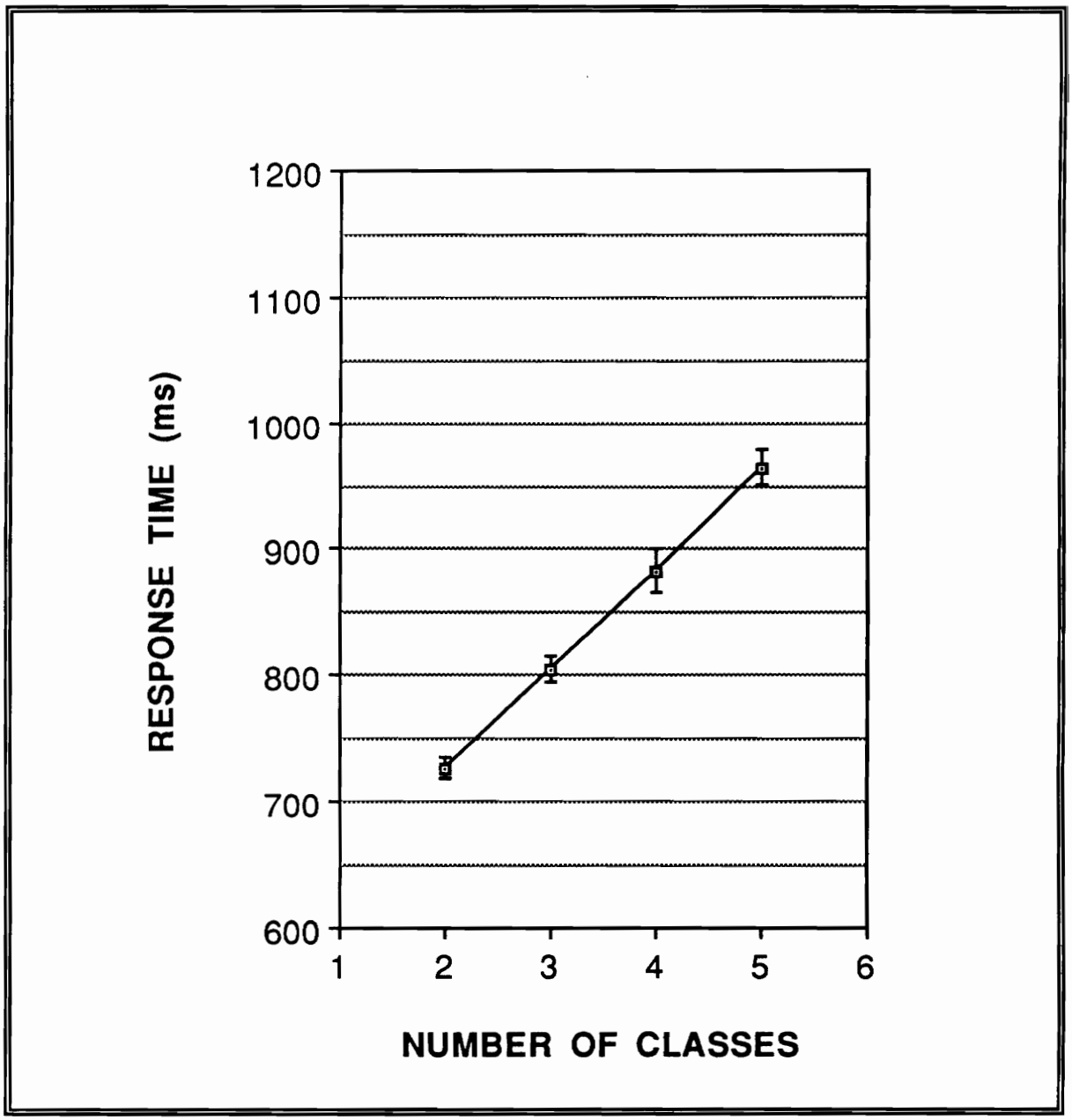


Figure 37. Effect of Number Of Classes, collapsed over subjects (N = 10), Blocks, Number Of Items, Presentation Order, Proximity, and Type, on response times in Experiment 1 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

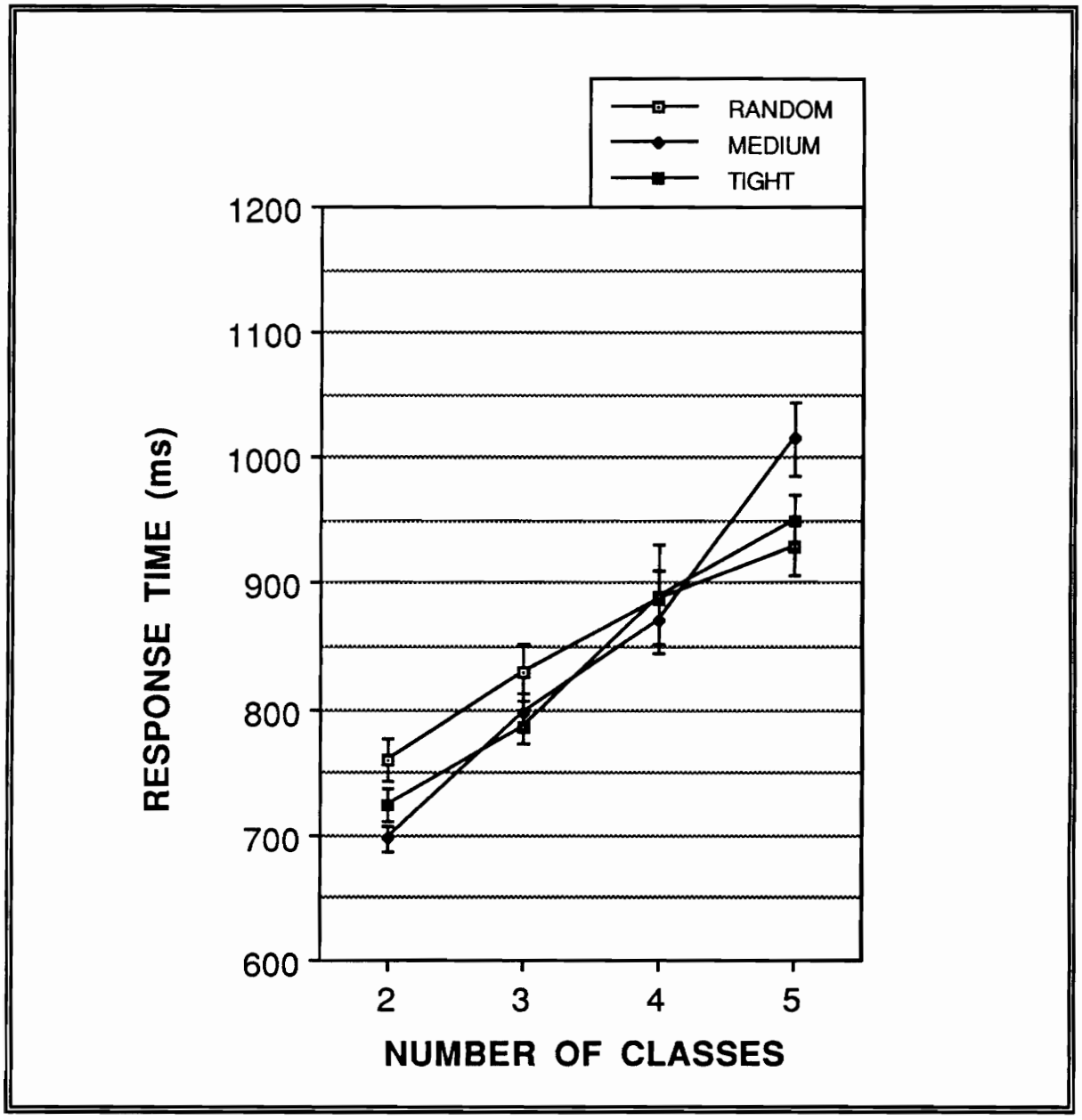


Figure 38. Two-factor interaction of Number Of Classes X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Items, Presentation Order, and Type, on response times in Experiment 1 ($p < 0.0232$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

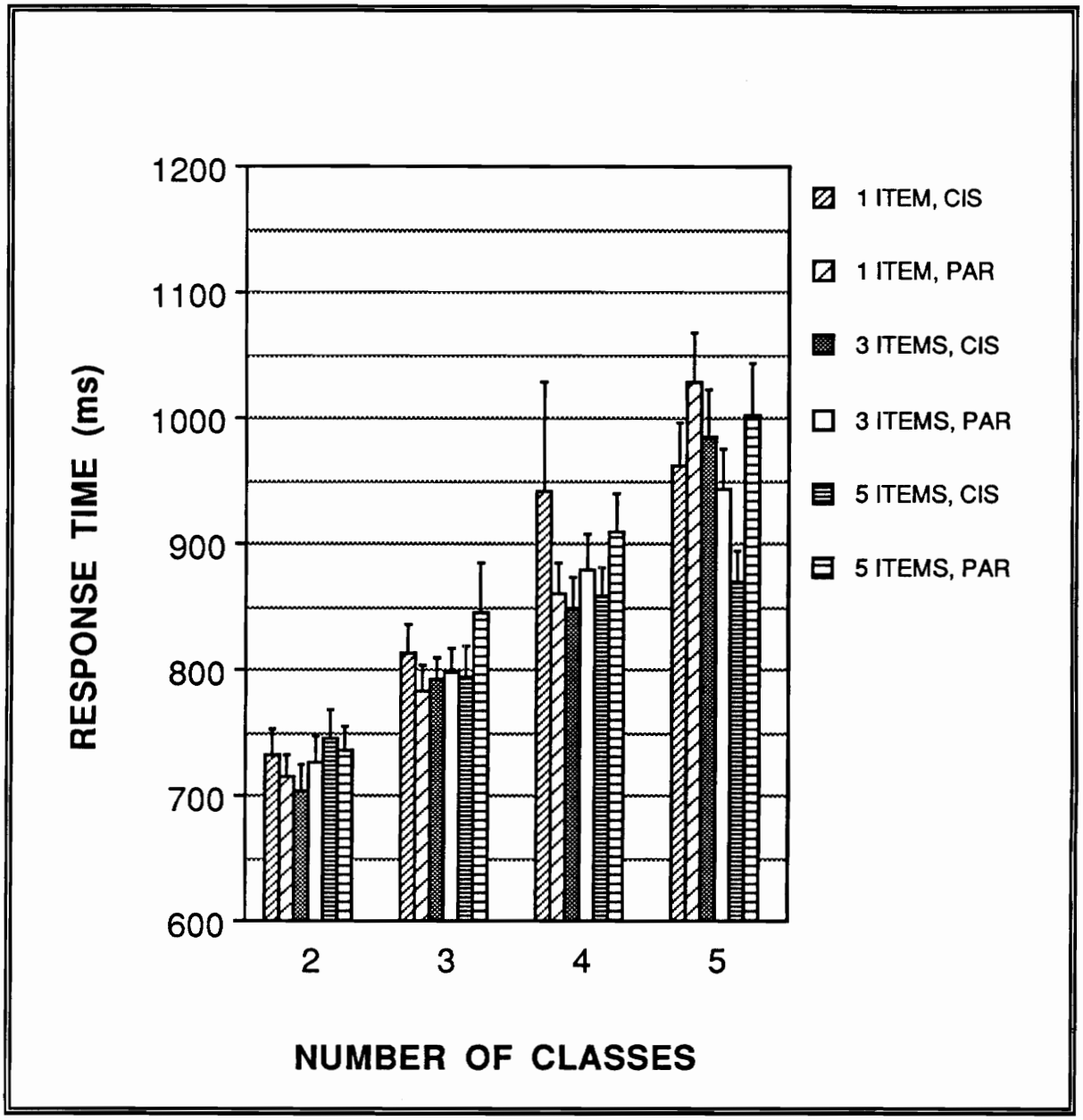


Figure 39. Three-factor interaction of Number Of Classes X Number Of Items X Presentation Order, collapsed over subjects (N = 10), Blocks, Proximity, and Type, on response times in Experiment 1 ($p < 0.0232$). Error bars indicate + 1 $s_{\bar{x}}$.

Items/Parallel), this interaction shows the limited influence of Number Of Items and display grouping on response times. As suggested above, it is expected that only Number Of Classes should affect the serial scanning of MSET items, and, thus, reaction times.

Effects due to Type. As illustrated in Figure 40, response time varies significantly with Type ($p < 0.0024$). Negative response trials require an average of 917.73 ms and positive response trials require an average of 771.85 ms. Also, the two-factor interaction between Number Of Classes and Type is significant ($p < 0.0404$). Visual inspection of Figure 41 allows confirmation of main effect findings that response times increase with negative trials and as Number Of Classes increases. The Number of Classes X Number Of Items X Type interaction is also significant ($p < 0.0409$) (Figure 42).

As suggested earlier, participants were unable to make positive identifications during negative probe trials. Therefore, STM scans tended to be exhaustive and, at times, followed by additional response choice processing.

Finally, the Number Of Classes X Number Of Items X Presentation Order X Proximity X Type interaction is significant ($p < 0.0130$).

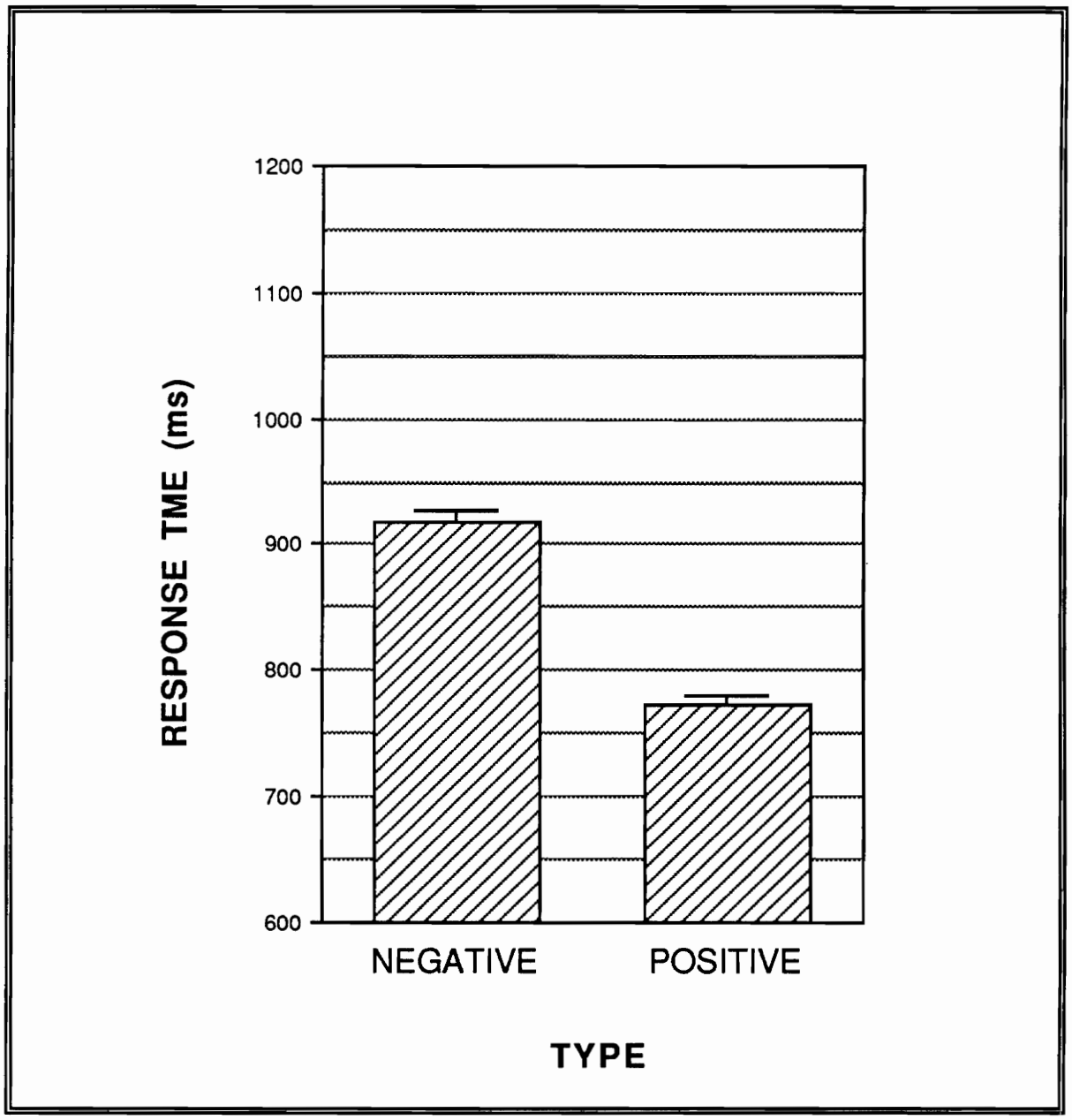


Figure 40. Effect of Type, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Presentation Order, and Proximity, on response times in Experiment 1 ($p < 0.0024$). Error bars indicate + 1 $s_{\bar{x}}$.

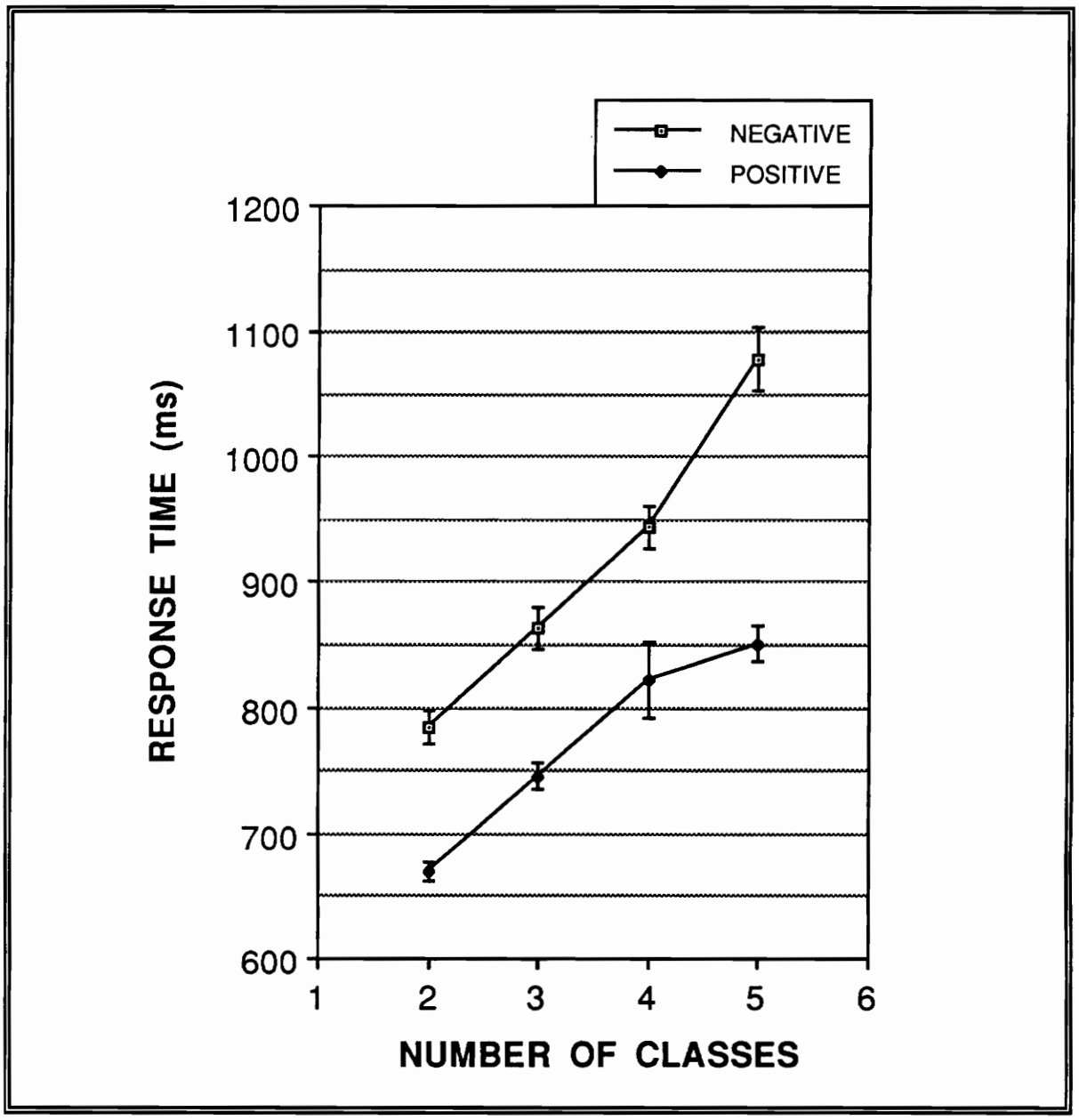


Figure 41. Two-factor interaction of Number Of Classes X Type, collapsed over subjects (N = 10), Blocks, Number Of Items, Presentation Order, and Proximity, on response times in Experiment 1 ($p < 0.0404$). Error bars indicate + 1 $s_{\bar{x}}$.

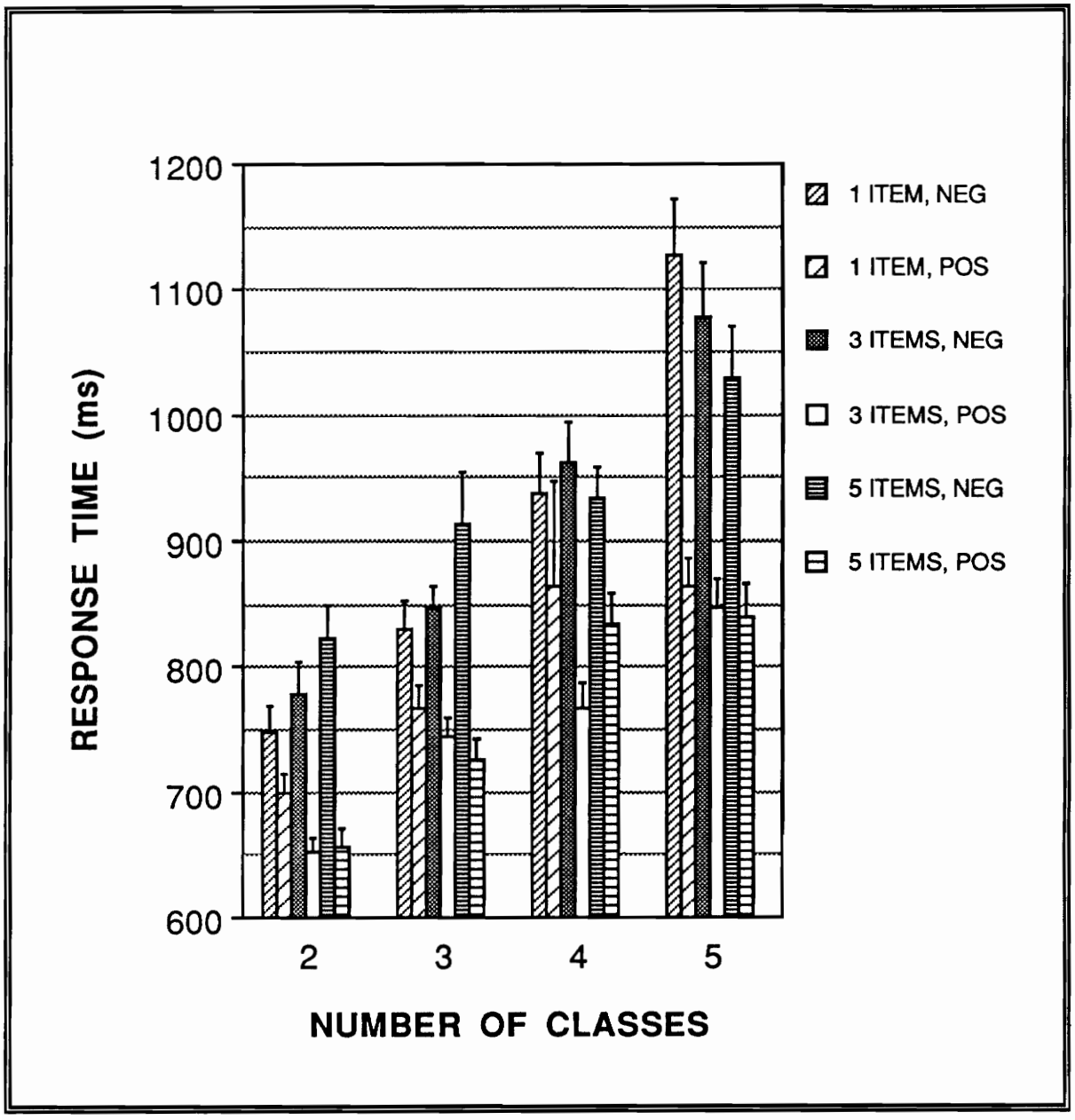


Figure 42. Three-factor interaction of Number Of Classes X Number Of Items X Type, collapsed over subjects (N = 10), Blocks, Presentation Order, and Proximity, on response times in Experiment 1 ($p < 0.0409$). Error bars indicate + 1 $s_{\bar{x}}$.

EXPERIMENT 2: METHOD

Experiment 2 was similar to Experiment 1 in that participants were required to view a display containing numerous symbols. In fact, all independent and dependent variables were identical across the experiments, except the independent variable of Probe Type. The chief difference was that Experiment 2 participants were required to store and recall symbol names (i.e., to parse the stimulus class chunks). It was hypothesized that display formatting methods which grouped same-class symbols would facilitate the recall of individual members from STM information chunks.

Participants

Ten male, undergraduate students served as participants. Students were recruited, screened, and paid in a manner identical to that used in Experiment 1. All participants passed the visual contrast sensitivity test.

Apparatus, Stimuli, and Experimental Environment

All equipment used was identical to that of Experiment 1.

Task

A modified Sternberg paradigm (Sternberg, 1966) was employed for the experiment. On each trial, 2 to 25 symbols were presented at various display locations. Participants were tasked to store the symbol names in STM. For example, if the screen showed the symbols: *frog*, *elephant*, *airplane*, and *helicopter*, the participant was required to store: *frog*, *elephant*, *airplane*, and *helicopter*. (In Experiment 1 participants stored only the class names).

Following the symbol display, the screen was blanked for 2 s and then a single probe symbol was presented at the center of the screen. The participant used the mouse buttons to respond "yes" or "no" to indicate whether the probe symbol had been presented during that trial. For example, given the symbols *frog*, *elephant*, *airplane*, and *helicopter* and the the probe symbol *frog*, the correct response is "yes" since the *frog* icon had been presented. Participants were instructed that speed and accuracy of response were important.

Design

As in Experiment 1, a five-factor, within-subjects, factorial design was employed to define the experimental trial conditions (Figure 5). The five factors were Number Of Classes (4 levels), Number Of Items per class (3 levels), Presentation Order (2 levels), Proximity (3 levels), and Probe Type (2 levels). Each participant viewed each of these 144 possible combinations in a unique randomized order. Each participant completed four blocks of trials, lasting about one hour each.

Number Of Classes (4 levels). Same as Experiment 1

Number Of Items (3 levels). Same as Experiment 1

Spatial Proximity (3 levels). Same as Experiment 1

Presentation Order (2 levels). Same as Experiment 1

Response Type (2 levels). The two levels of Type describe whether the correct answer for a trial was "yes" or "no". In the "yes" case, the probe symbol was a random symbol that had been displayed during the symbol presentation phase.

In the "no" case, the probe symbol was randomly drawn from those symbols which had not been displayed during the symbol presentation phase, but belonged to a symbol class which had been displayed. For example, if the symbols *frog*, *elephant*, *airplane*, and

helicopter had been displayed during the symbol presentation phase, *dog* or *tank* would have been acceptable probes.

Blocks (3 levels). Same as Experiment 1

Dependent Measures

Same as Experiment 1

Procedure

The procedure was the same as that of Experiment 1. See Appendix E for task instructions to participants.

EXPERIMENT 2: RESULTS AND DISCUSSION

Due to the large volume of data, it is useful to consider the process or strategy participants might have used to solve each problem. During a trial, it is suggested that a participant might:

- o glance at the *bear* symbol and use some strategy to encode this information into STM.
- o glance at the *dog* symbol and use some strategy to encode this information into STM.
- o etc. (It is likely that in some instances the presentation set disappeared before the participant had time to view all of the symbols.)
- o After symbol presentation, the probe was presented and the participant scanned the items stored in his STM.

Some insights into participants' problem solving strategies were gained during *post hoc* debriefing interviews: first, participants attempted to chunk items. That is, they tried to view and encode items belonging to one class before shifting to items belonging to other classes. For example, participants preferred to encode *bear*, *dog*, then *fork*, *spoon* as opposed to a serial encoding of bear, *fork*, *dog*, *spoon*. Presumably, this strategy allowed participants to create MSETs in which items were chunked into same-class strings. Even in cases where same-class symbols were not proximal, participants searched across the display, incurring a time cost, in order to sequentially encode same-class (i.e., easily "chunkable") symbols.

Second, most participants attempted to use a chunking shorthand with some of the symbols. For example, the digits *1*, *3*, *5* might have been encoded as "odd", the symbols

cup and *pitcher* might have been encoded as handle (both had handles), and the symbols *airplane*, *air-to-air missile*, and *radar* might have been encoded as "air".

Third, some participants noted storing information in different ways. They described storing some information as words and other information as mental images. This strategy is consistent with Wickens' (1984) multiple resource theory. Specifically, when phonetic STM was overloaded, participants shifted some of the task load to visual STM.

Data Analysis Overview

As in Experiment 1, there were 432 data observations for each dependent measure (3 blocks x 144 trials per block) across the factorial combinations of Blocks (3 levels), Number Of Classes (4 levels), Number Of Items (3 levels), Presentation Order (2 levels), Proximity (2 levels), and Type (2 levels). Thus, the 10 participants provided a total of 4320 data points for each dependent measure. The data for each dependent measure (i.e., SWAT, reaction time, and error rate) were subjected to a five-factor, within-subjects ANOVA procedure. Data were averaged across blocks during the ANOVA. Complete tabulations of the ANOVA results are listed in Tables H1 through H3. The findings of each ANOVA are discussed below.

SWAT Ratings

As in Experiment 1, a SWAT scale development procedure was completed by each participant prior to data collection. Each participant sorted 27 cards representing low to high workload situations. Scale values for each individual were calculated using the additive polynomial model incorporated in the SWAT conjoint analysis software (Reid et al., 1989).

During scale development, there were no independence violations and six joint independence violations across the card sorts for all 10 participants. This relatively low number of violations fell well below the suggested maximum limit of 20 independence and 20 joint independence violations (Reid et al., 1989) and indicated that each participant's card sort was internally valid. Therefore, no participants were required to repeat the scale development procedure.

Kendall's W, the Coefficient of Concordance, across participants was 0.88. This exceeded the suggested minimum of 0.75 (Reid et al., 1989) and indicated that participants provided homogeneous card sorts. Accordingly, a single scaling solution was used to represent their composite view of workload.

The previously developed scaling values were used to transform the three dimension ratings (i.e., *time*, *mental effort*, *stress*) from each trial into a single overall SWAT rating ranging between 0 and 100. A scaling value of 100 represented the highest possible workload. The average SWAT rating across all 4320 trials was 34.42.

Effects due to display complexity. In general, SWAT ratings increase as the number of symbols increases. As illustrated in Figure 43, SWAT ratings increase monotonically as the Number Of Classes increases ($p < 0.0001$). A Newman-Keuls test (Table H4) indicates that all levels of Number Of Classes differ significantly ($p < 0.05$). As shown in Figure 44, SWAT ratings also increase monotonically with the Number Of Items ($p < 0.0001$). A Newman-Keuls test (Table H5) indicated that all levels of Number Of Items differ significantly ($p < 0.05$).

The Number Of Classes by Number Of Items two-factor interaction also is significant ($p < 0.0001$). Visual inspection of Figure 45 allows that SWAT ratings increase more rapidly when the Number Of Items equals one than when the Number Of Items equals three or five.

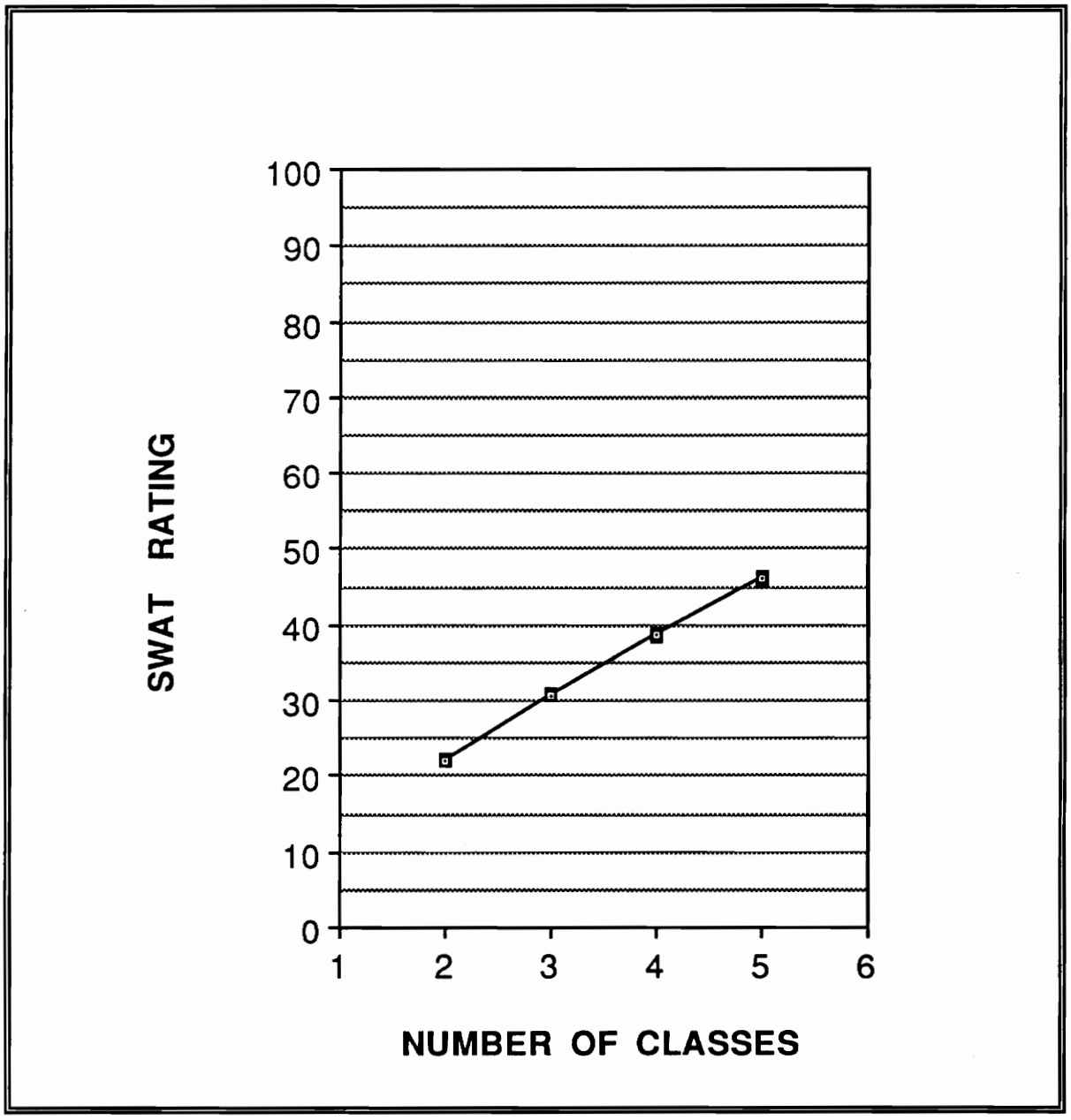


Figure 43. Effect of Number Of Classes, collapsed over subjects (N = 10), Blocks, Number Of Items, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 2 ($p < 0.0001$). Error bars (not visible due to small size) indicate $\pm 1\ s_{\bar{x}}$.

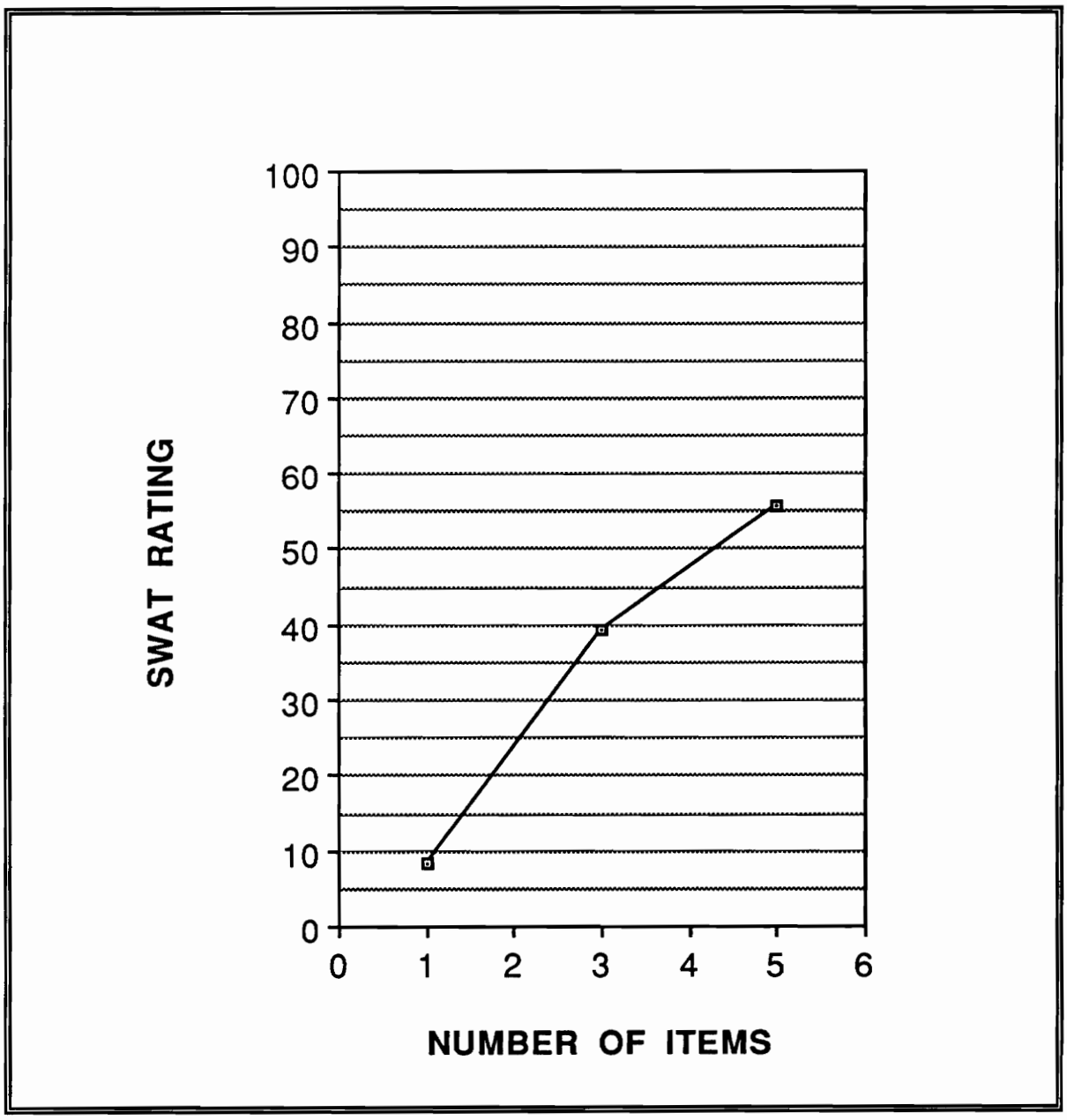


Figure 44. Effect of Number Of Items, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 2 ($p < 0.0001$). Error bars (not visible due to small size) indicate $\pm 1\ s_{\bar{x}}$.

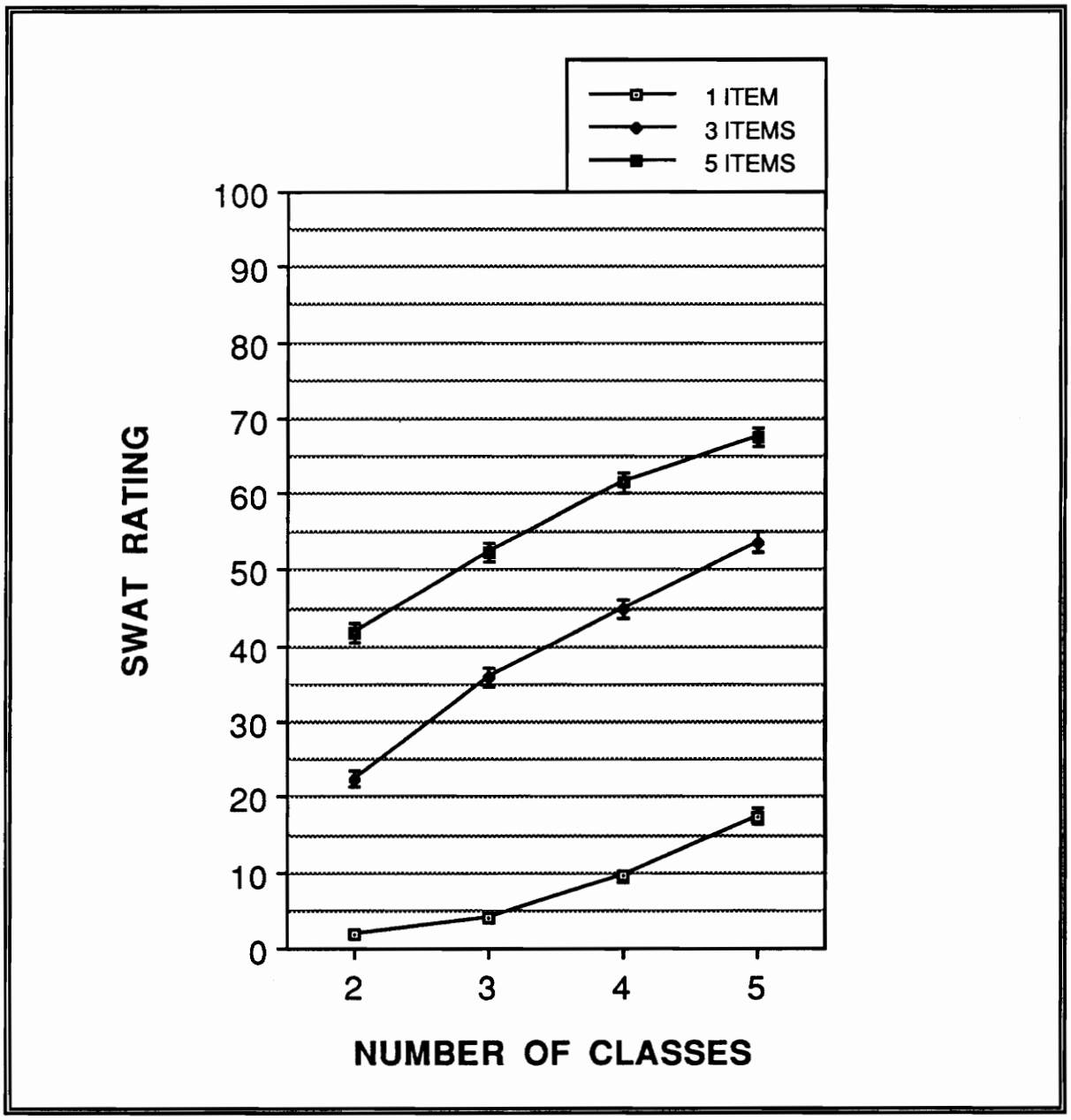


Figure 45. Two-factor interaction of Number Of Classes X Number Of Items, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 2 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

An auxiliary analysis was performed on the total number of symbols (Number Of Classes x Number Of Items) displayed on the screen. As shown in Figure 46 (and Table H6), SWAT ratings increase monotonically with the number of symbols. Viewing Table H6, it is especially useful to compare the 3-Classes/5-Items combination to the 5-Classes/3-Items combination. Each combination provides a total of 15 symbols and SWAT ratings do not differ ($p > 0.05$).

Returning to Figure 45, it can be reasoned that when the Number Of Items equals one, each increment in class size generates one additional symbol. When the Number Of Items equals three or five, each increment in class size generates two additional symbols. Therefore, the Number Of Classes X Number Of Items interaction is consistent with the argument that total number of symbols dictates task difficulty rather than the independent effects due to Number Of Classes and Number Of Items.

It is interesting that Experiment 1 task difficulty was driven by unique combinations of Number Of Classes (i.e., number of chunks) and Number Of Items (i.e., number of items per chunk), while Experiment 2 task difficulty was driven by the total number of symbols. Considering participants' task operations for these experiments, it is clear that Experiment 1 required participants to remember chunk labels. Experiment 2 required participants to remember (i.e., create an MSET of) all symbols displayed. Furthermore, it seems that specific Number-Of-Classes/Number-Of-Items combinations did not impact the subjective difficulty of this task.

Effects due to display grouping. As in Experiment 1, display formats which group same-class symbols are associated with decreased subjective workload. For example, the CIS Order was judged to require less workload than the Parallel Order ($p < 0.0014$). As illustrated in Figure 47, average SWAT ratings are 30.26 for the CIS Order and 38.56 for the Parallel Order. The main effect of Proximity also is significant ($p < 0.0001$). As

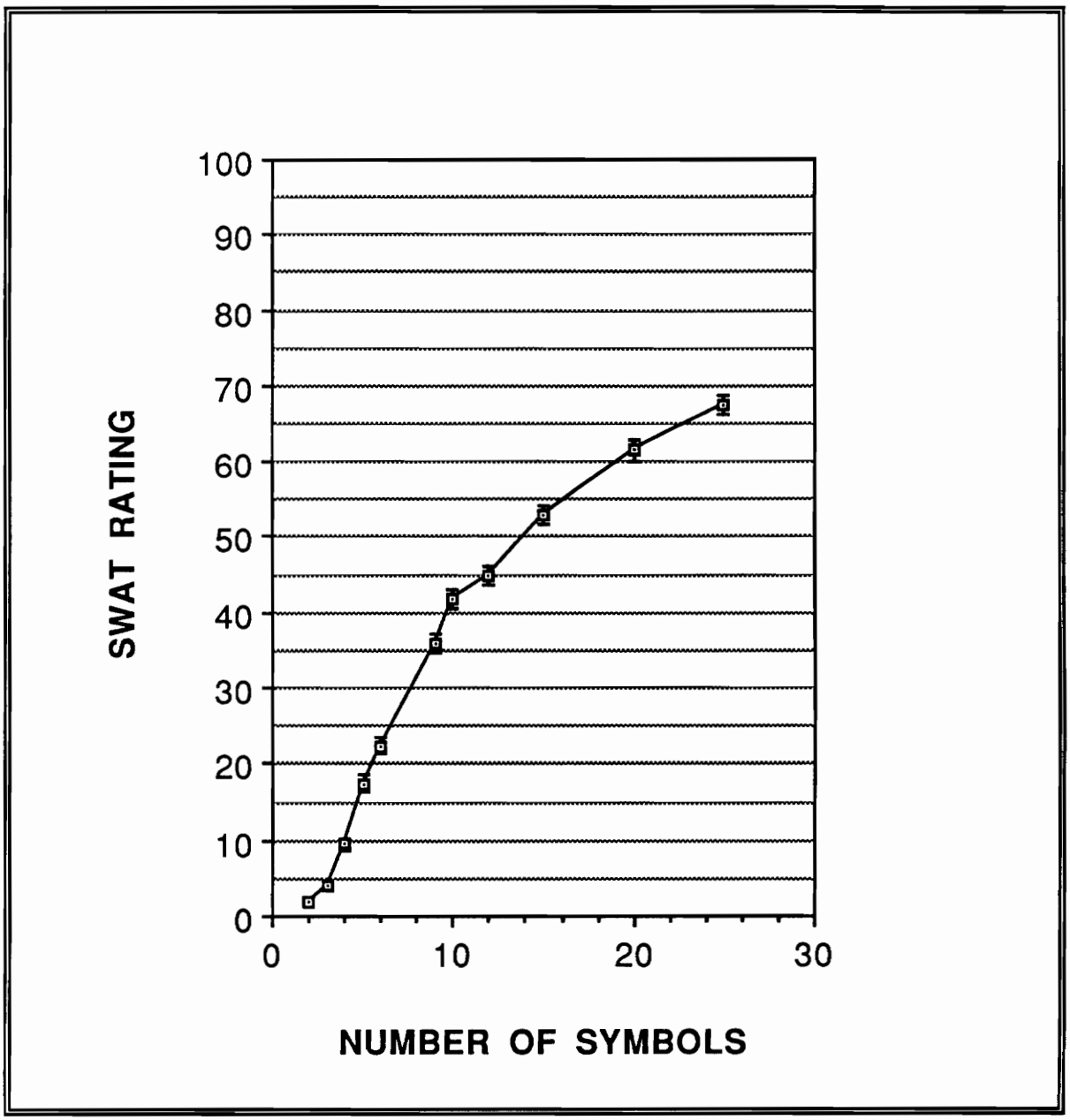


Figure 46. The effect of the number of symbols, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on SWAT ratings in Experiment 2. Error bars indicate $\pm 1\ s_{\bar{x}}$. Note: There were two cases in which the number of symbols equalled 15. Average means and $s_{\bar{x}}$ s are shown.

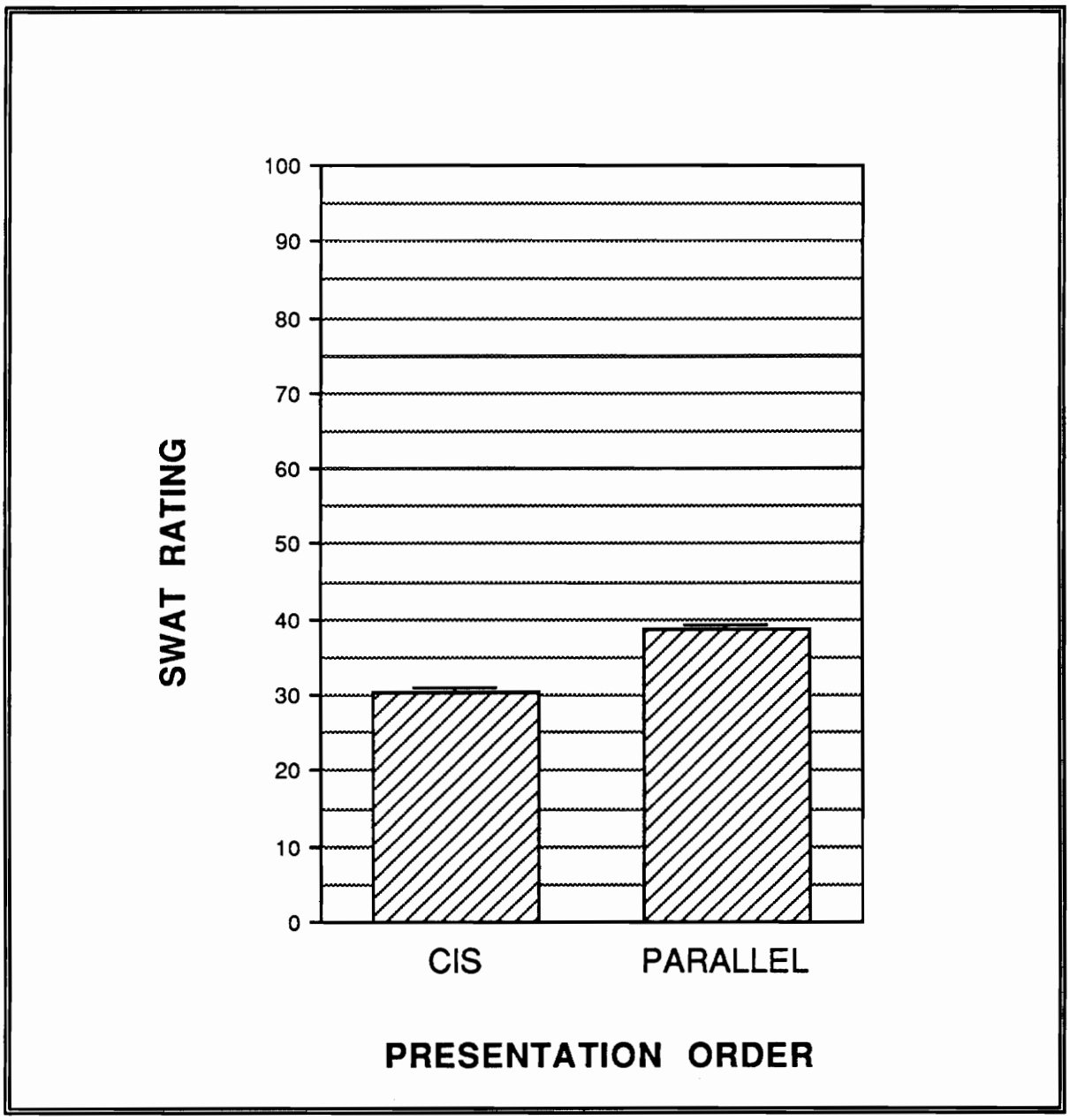


Figure 47. Effect of Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Proximity, and Type, on SWAT ratings in Experiment 2 ($p < 0.0014$). Error bars indicate + 1 $s_{\bar{x}}$.

shown in Figure 48, SWAT ratings decrease as Spatial Proximity increases. A Newman-Keuls test (Table H7) indicates that subjective workload differs at each level of Proximity ($p < 0.05$).

Finally, the interaction between Presentation Order and Proximity is significant ($p < 0.0001$). In Figure 49 (and Table H14), it can be seen that Proximity does not affect SWAT ratings under the CIS Order, but that SWAT ratings decrease as Proximity increases under the Parallel Order. The Parallel-Order/Tight-Proximity condition provided SWAT ratings equivalent to the CIS presentations.

In summary, display grouping formats reduced subjective workload because they helped individuals view, encode, and store same-class (i.e., easily chunkable) symbols in a sequential manner.

However, the SWAT data associated with the Tight and Medium Proximity levels are not consistent with participants' comments. During debriefing interviews, some participants reported that they preferred the Medium Proximity condition, to the Tight Proximity condition, and, finally, the Random Proximity condition. They explained that Tight Proximity symbols were too close together and therefore hard to differentiate from other symbols. These *post hoc* verbal reports suggest a proximity limit which should not be exceeded in practical display design.

Combined effects of complexity and grouping. As in Experiment 1, the subjective workload reduction associated with display grouping formats (i.e., CIS Order, Tight Proximity) becomes pronounced as display complexity (i.e., the number of symbols) increases. For example, the interaction between Number Of Classes and Presentation Order is significant ($p < 0.0002$). As shown in Figure 50 (and Table H9) SWAT ratings are lower for CIS Order across all levels of Number Of Classes ($p < 0.001$), yet the significance of the Number Of Classes X Presentation Order interaction indicates that

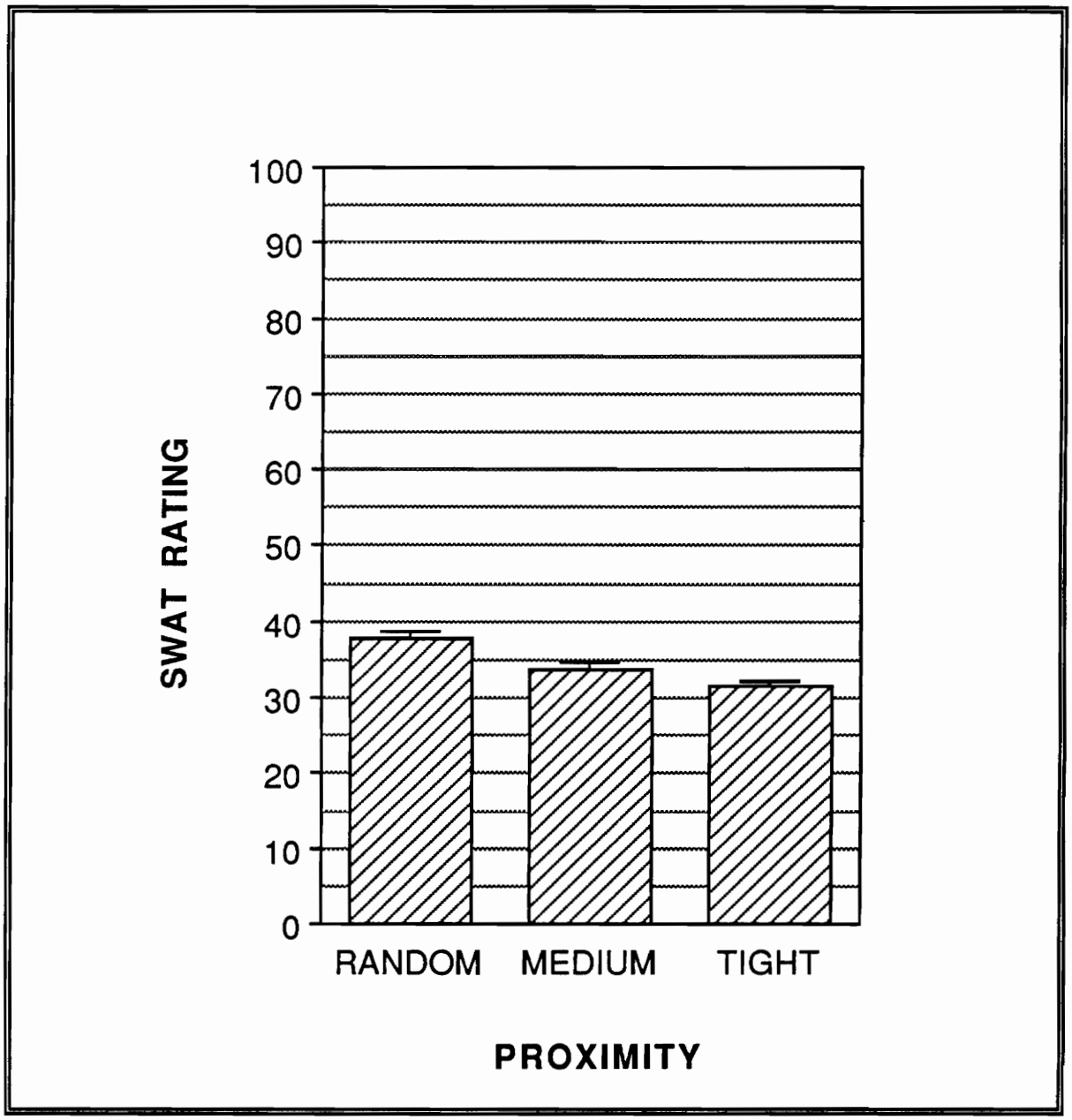


Figure 48. Effect of Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Presentation Order, and Type, on SWAT ratings in Experiment 2 ($p < 0.0001$). Error bars indicate + 1 $s_{\bar{x}}$.

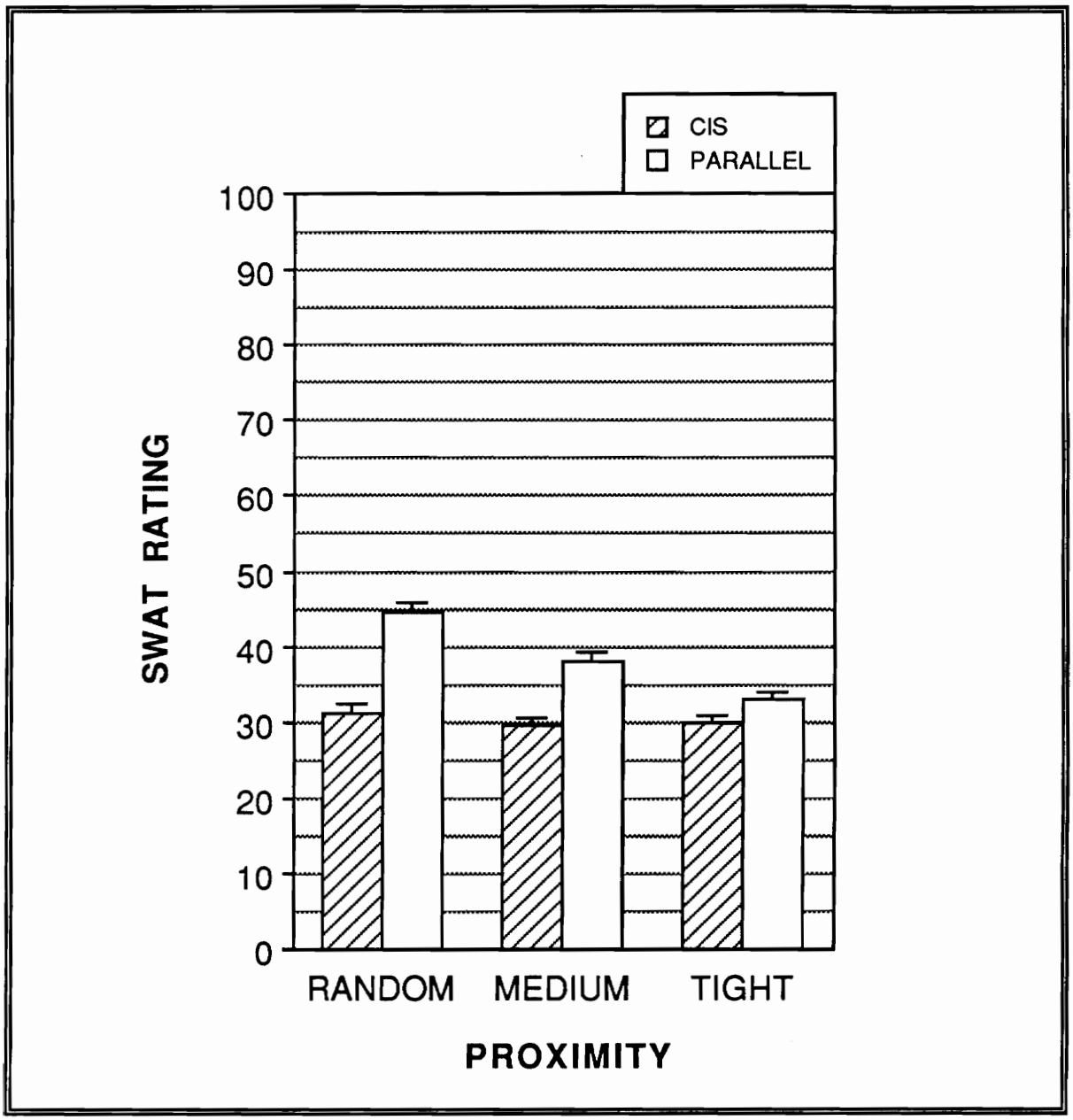


Figure 49. Two-factor interaction of Presentation Order X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, and Type, on SWAT ratings in Experiment 2 ($p < 0.0001$). Error bars indicate + 1 $s_{\bar{x}}$.

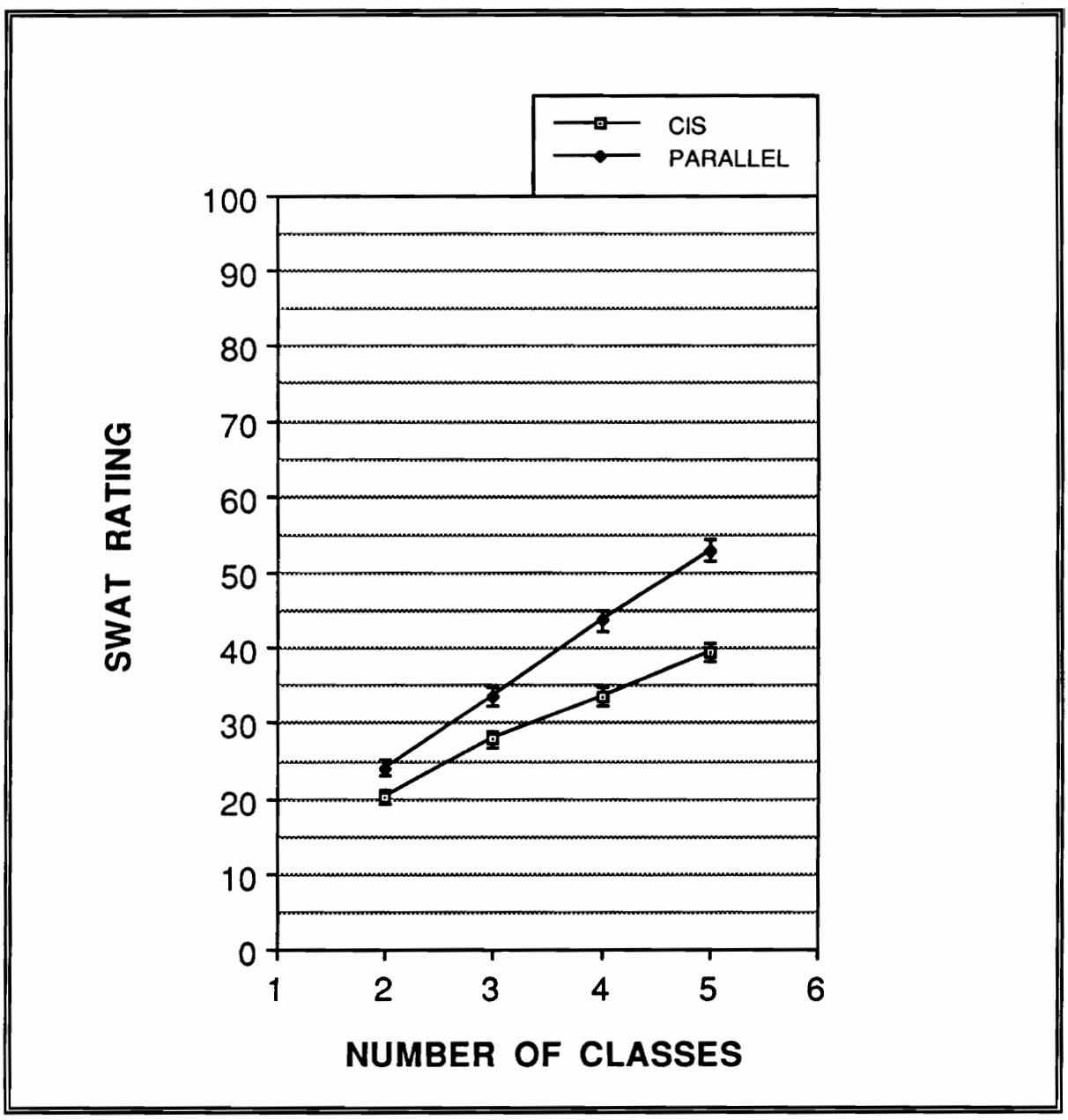


Figure 50. Two-factor interaction of Number Of Classes X Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Items, Proximity, and Type, on SWAT ratings in Experiment 2 ($p < 0.0002$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

the SWAT reduction benefit of the CIS Order becomes more pronounced as the Number Of Classes increases.

As shown in Figure 51, the interaction between Number Of Items and Presentation Order also was significant ($p < 0.0085$). Simple-effect F-tests (Table H10) indicated Presentation Order did not significantly effect SWAT ratings when the Number Of Items equaled one ($p > 0.05$), but did when the Number Of Items equaled three or five ($p < 0.001$). As in Experiment 1, it is noted that when the Number Of Items equaled one, participants saw between 2 and 5 symbols (classes), presented either simultaneously (i.e., Parallel Order) or sequentially (i.e., CIS Order). In this case, each symbol represented a single class and, therefore, facilitative grouping provided no additional organization. However, when the number of items was three or five, participants found CIS Order helped them organize information before storing it as an MSET string.

The interaction between Number Of Items and Proximity is significant ($p < 0.0001$). In Figure 52 (and Table H 11), it can be seen that Proximity has no effect on SWAT ratings when the Number Of Items equaled one. As the Number Of Items increases, the workload associated with low proximity increases more rapidly than that associated with higher proximity groupings. As in Experiment 1, when the Number Of Items equaled one, there was only one symbol per class, so it is not meaningful to think of a symbol as being in random, medium, or tight Proximity to other members of its class. Therefore, it is appropriate that there is no significant effect of Proximity when the Number Of Items equals one. However, when the Number Of Items equals three or five, subjective workload is reduced as Proximity increases from random, to medium, to tight. Again, it appears that increased Proximity helped participants organize information before storing it as an MSET string.

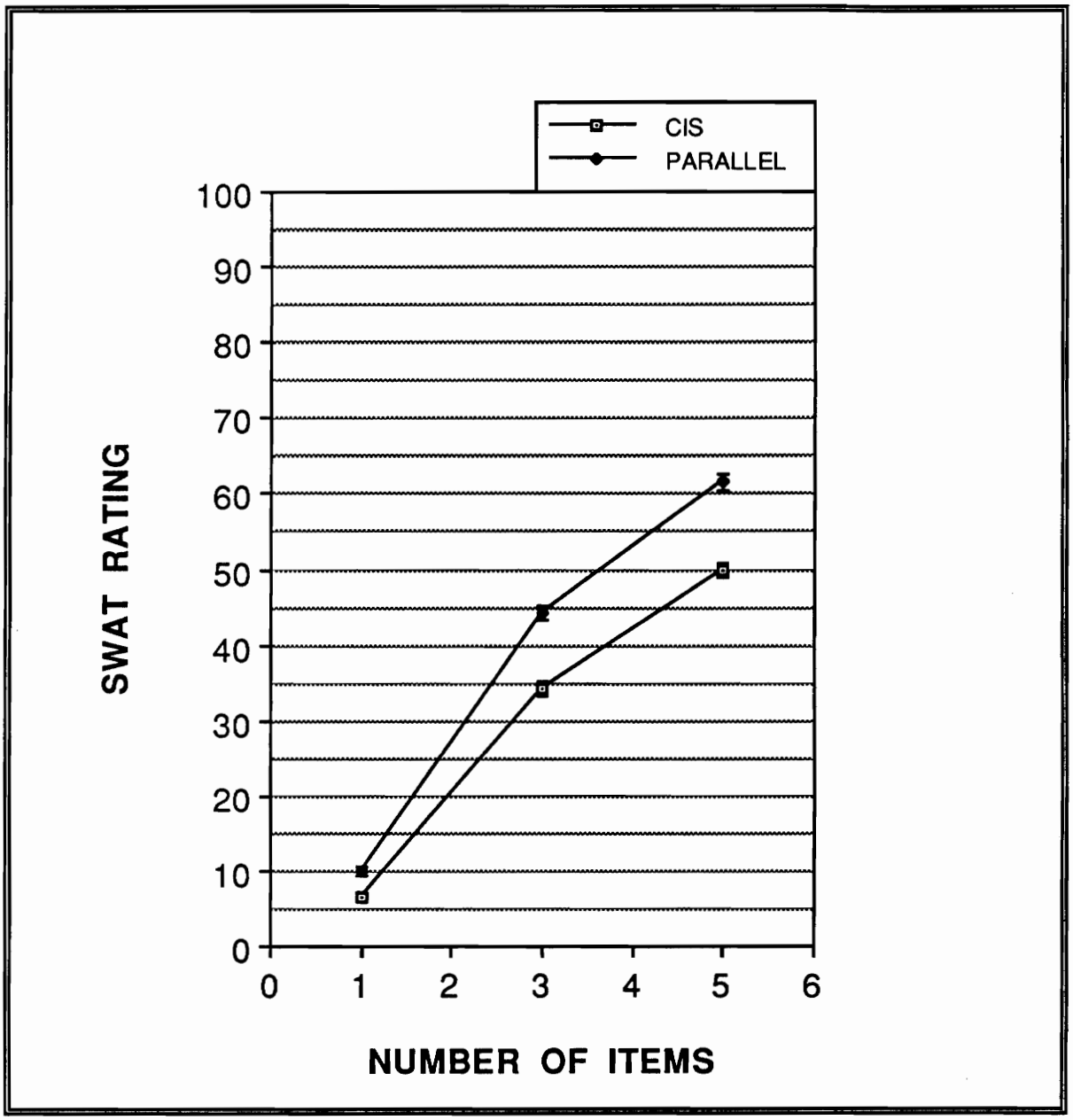


Figure 51. Two-factor interaction of Number Of Items X Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Classes, Proximity, and Type, on SWAT ratings in Experiment 2 ($p < 0.0085$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

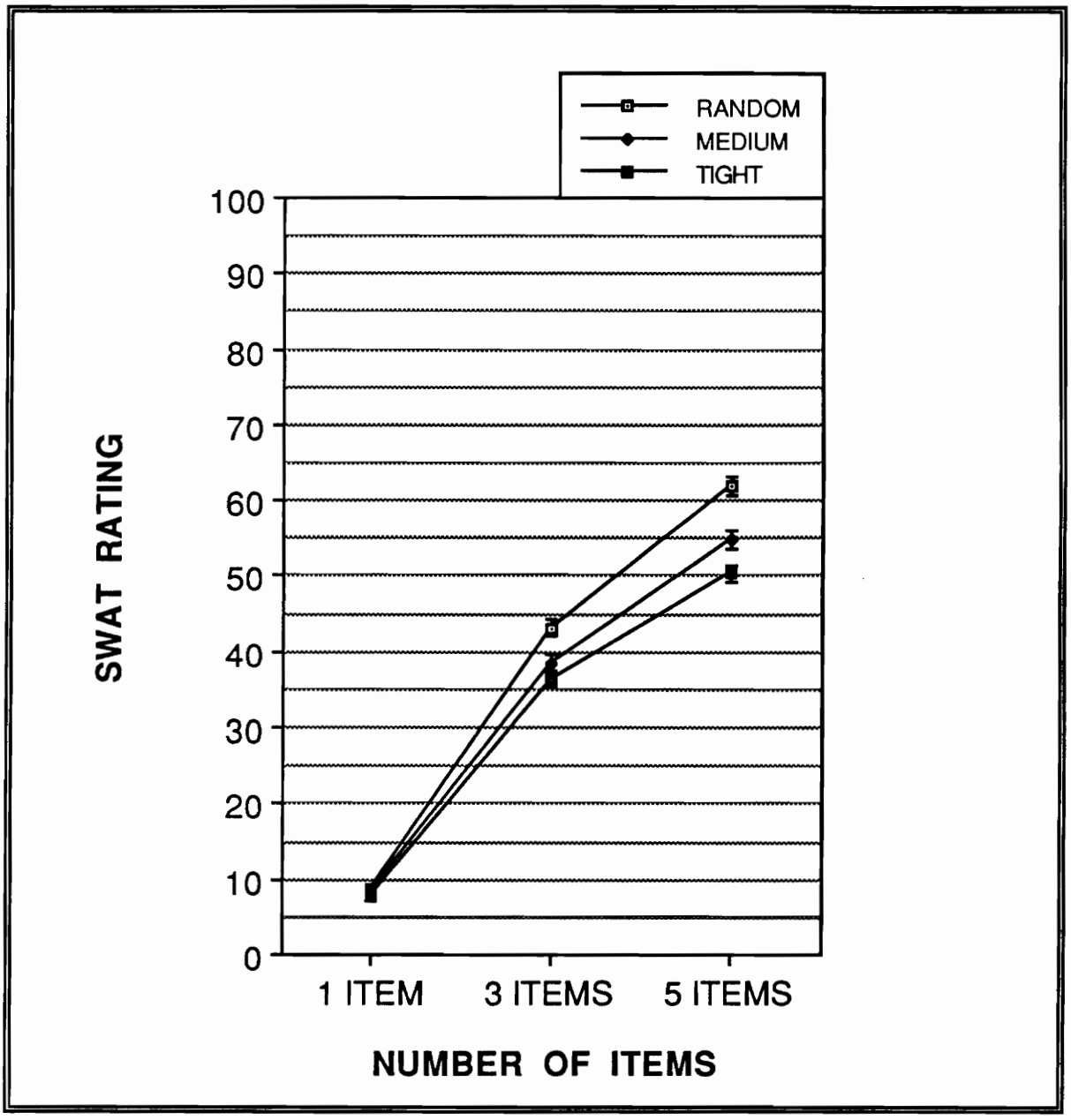


Figure 52. Two-factor interaction of Number Of Items X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, and Type, on SWAT ratings in Experiment 2 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

As shown in Figure 53, the Number Of Items X Presentation Order X Proximity interaction also is significant ($p < 0.0019$). In general, this three-factor interaction is consistent with its component two-factor interactions. For example, a series of simple-effect F-tests (Table H12) indicates that the simple two-factor interaction of Presentation Order X Proximity is not significant when the Number Of Items equals one, but that it is significant when the Number Of Items equals three or five. A Newman-Keuls test (Table H13) indicates that Figure 53a is consistent with the Number Of Items X Presentation Order interaction and the Number Of Items X Proximity interaction. Specifically, when the Number Of Items equals one, SWAT ratings do not vary significantly with Presentation Order or Proximity.

Figures 53b and 53c (and Tables H14 and H15) are consistent with the Presentation Order X Proximity interaction. Specifically, under the CIS Presentation Order, SWAT ratings do not vary significantly with Proximity level, but SWAT ratings increase as Proximity decreases under the Parallel Order conditions. The Parallel-Order/Tight-Proximity condition provides SWAT ratings similar to CIS presentations.

Effects due to Type. The main effect of Type is significant ($p < 0.0140$). As illustrated in Figure 54, average SWAT ratings are 35.21 for positive trials and 33.62 for negative trials. Again, it is suggested that in Positive Type trials, participants often matched an MSET item to the probe. A positive match was not possible in negative trials and participants felt less confident of their responses.

The four-factor interaction among Number Of Items, Presentation Order, Proximity, and Type is significant ($p < 0.0634$). This interaction stems from a small number of combinations (i.e., 4 out of 18) of Number Of Items, Presentation Order, and Proximity in which Positive Type trials were rated as more difficult than Negative Type trials. There are no other discernable patterns in this interaction effect.

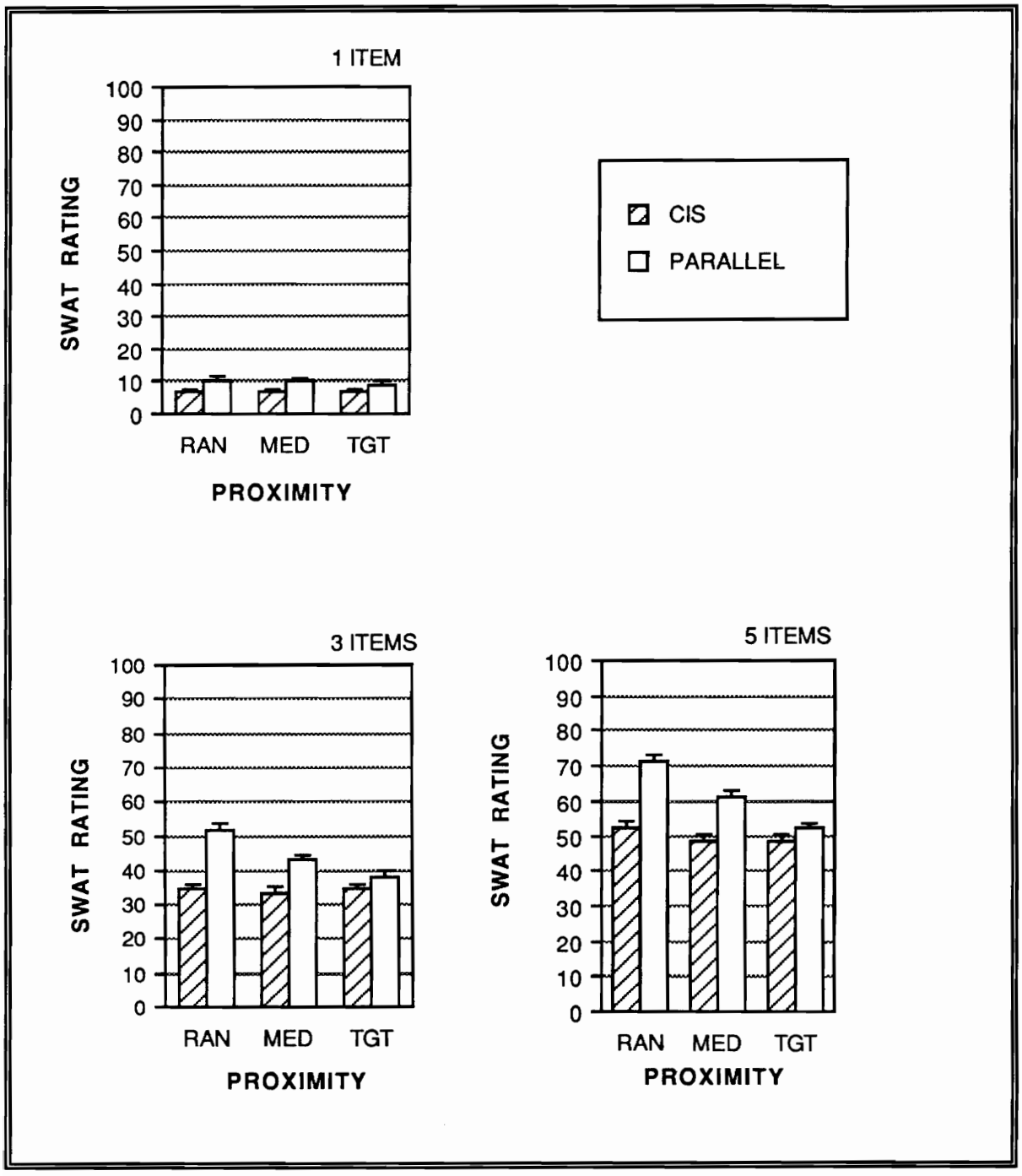


Figure 53. Three-factor interaction of Number Of Items X Presentation Order X Proximity, collapsed over subjects (N = 10), Blocks, Number Of Classes, and Type, on SWAT ratings in Experiment 2 ($p < 0.0019$). Error bars indicate + 1 $s_{\bar{x}}$.

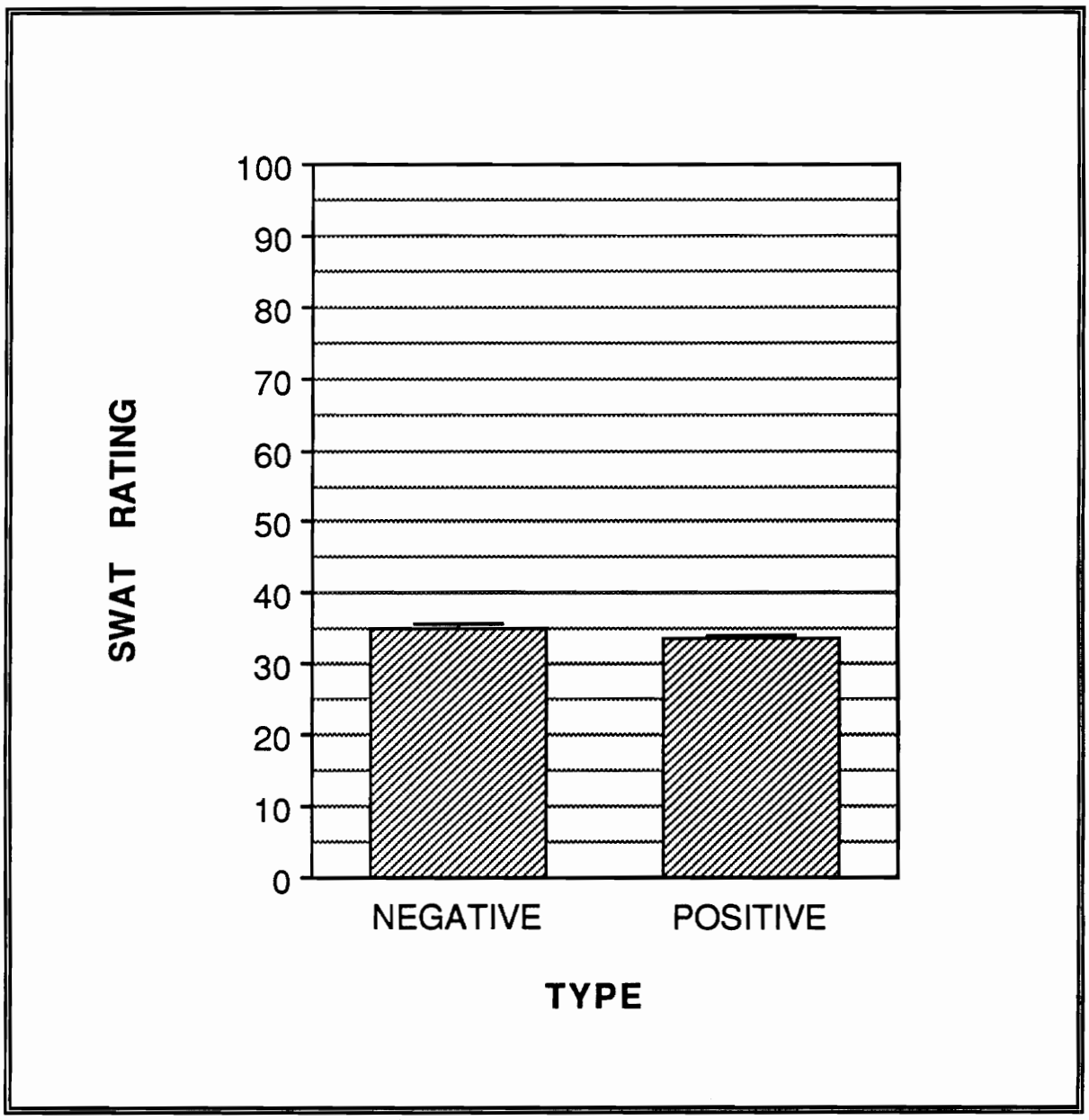


Figure 54. Effect of Type, collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Presentation Order, and Proximity, on SWAT rating in Experiment 2 ($p < 0.0140$). Error bars indicate + 1 $s_{\bar{x}}$.

Error Rate

Over all 4320 trials, participants made 472 errors (10.9 percent average error rate).

Effects due to display complexity. Display complexity factors affected error rate in a manner similar to their effects on subjective workload. That is, error rates increase with the total number of displayed symbols.

As shown in Figure 55 (and Tables H16), error rate increases with increasing Number Of Classes ($p < 0.0001$). A Newman-Keuls test (Table H16) indicates that error rates differ between the extreme levels of Number Of Classes, but not between the mid-levels of this factor. Similarly, as shown in Figure 56, error rate increases with increasing Number Of Items ($p < 0.0001$). A Newman-Keuls test (Table H17) indicates that error rates differ between all item levels ($p < 0.05$).

The interaction of Number Of Classes and Number Of Items is significant ($p < 0.0268$). Viewing Figure 57, it appears that error rate increases less rapidly when Number Of Items equals one than when the Number Of Items equals three or five.

As an auxiliary analysis, the effect of the total number of symbols (i.e., Number Of Classes x Number Of Items) displayed during any trial is shown in Figure 58. It can be seen in Figure 58 (and Table H18) that error rate tends to increase with the total number of symbols. In Table H18, a comparison of the 3-Classes/5-Items combination to the 5-Classes/3-Items combination (where the number of symbols equals 15) indicates that error rates do not differ significantly.

The significant Number Of Classes X Number Of Items interaction is consistent with the conclusion that the total number of symbols, is the best indicant of task difficulty.

Effects due to display grouping. Error rates do not differ significantly between the CIS and the Parallel Presentation Orders ($p = 0.5380$), where the average error rates were 0.113 errors and 0.106, respectively. Similarly, error rates do not vary significantly among

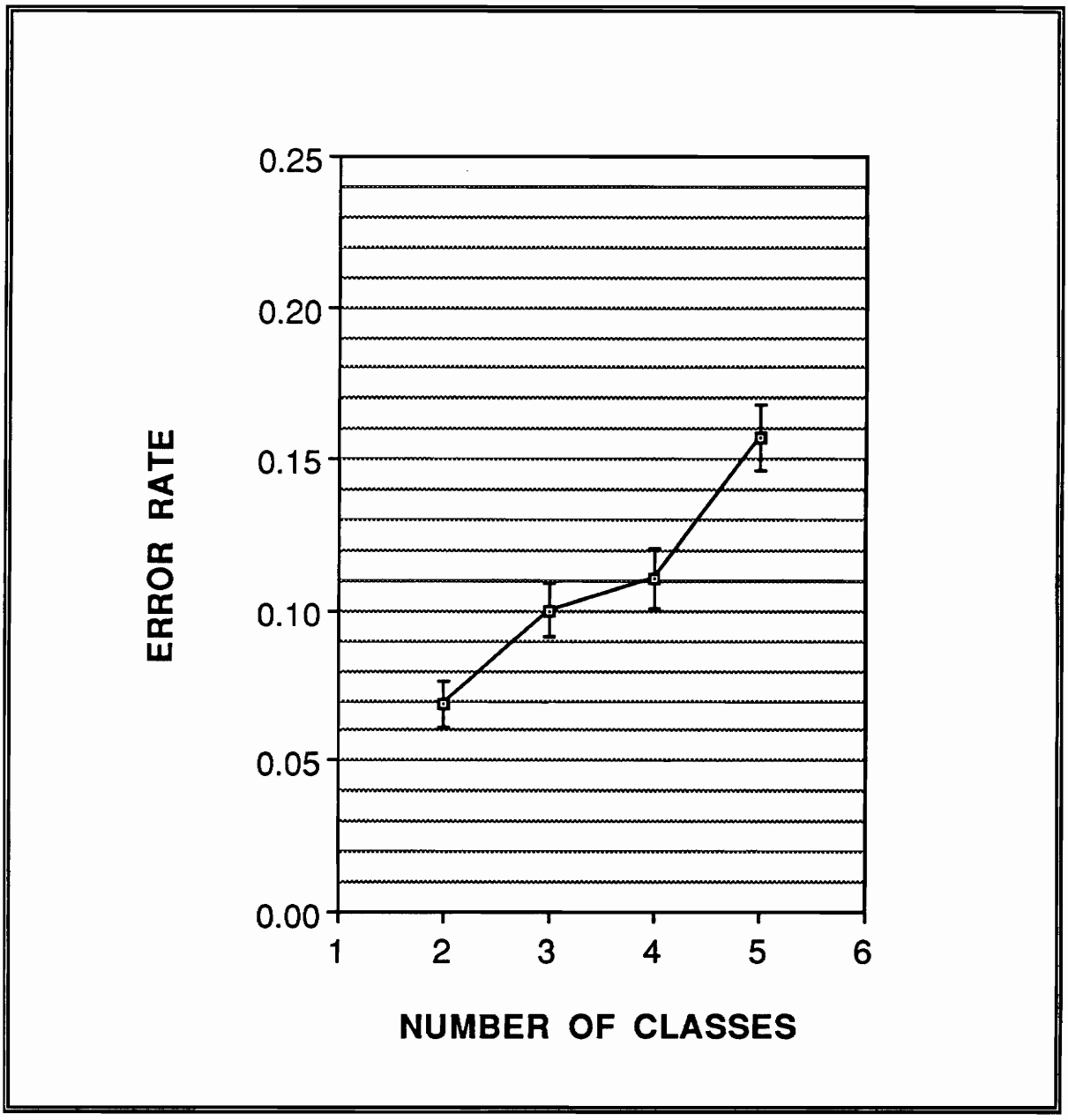


Figure 55. Effect of Number Of Classes, collapsed over subjects (N = 10), Blocks, Number Of Items, Presentation Order, Proximity, and Type, on error rates in Experiment 2 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

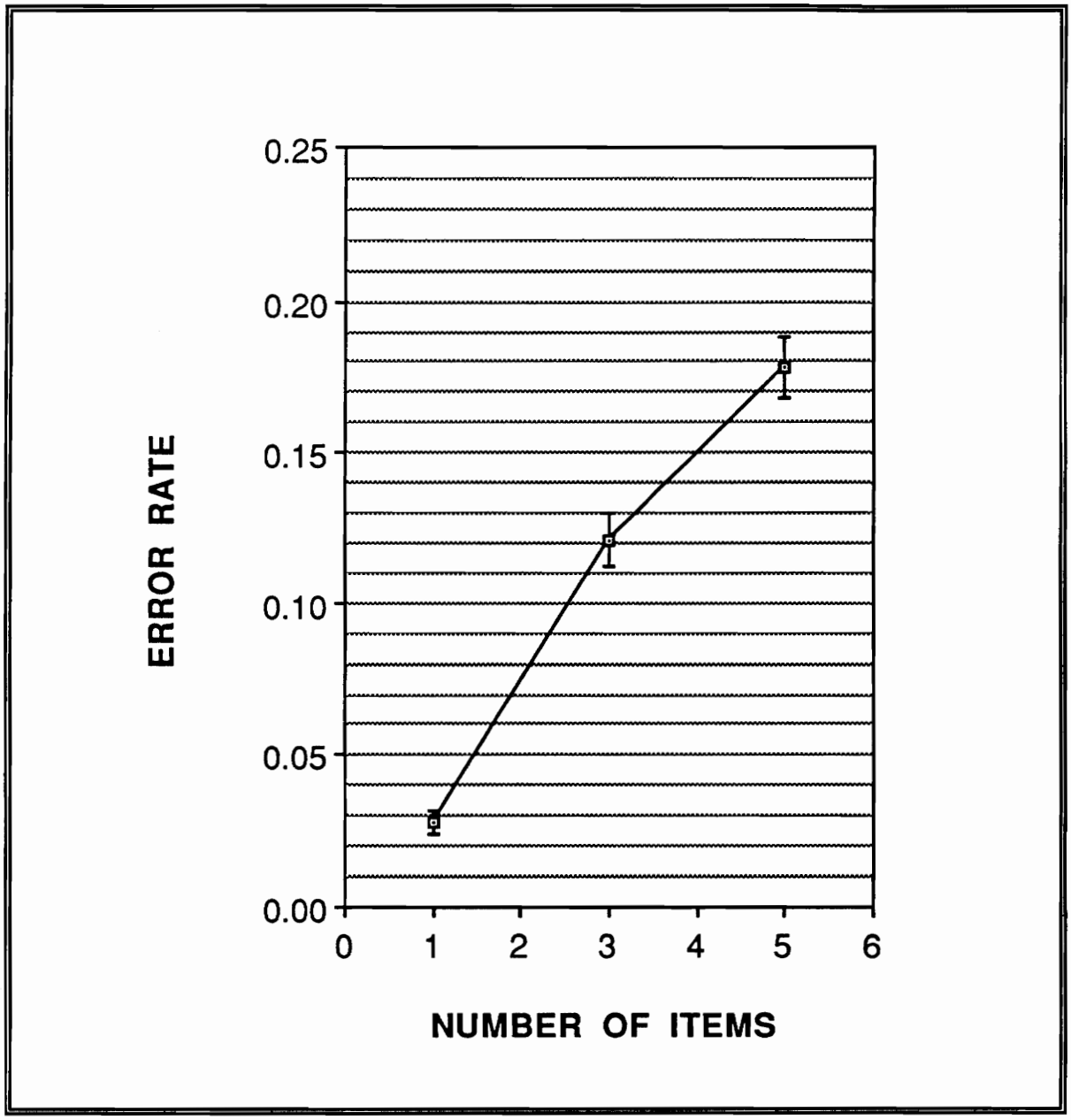


Figure 56. Effect of Number Of Items, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, Proximity, and Type, on error rates in Experiment 2 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

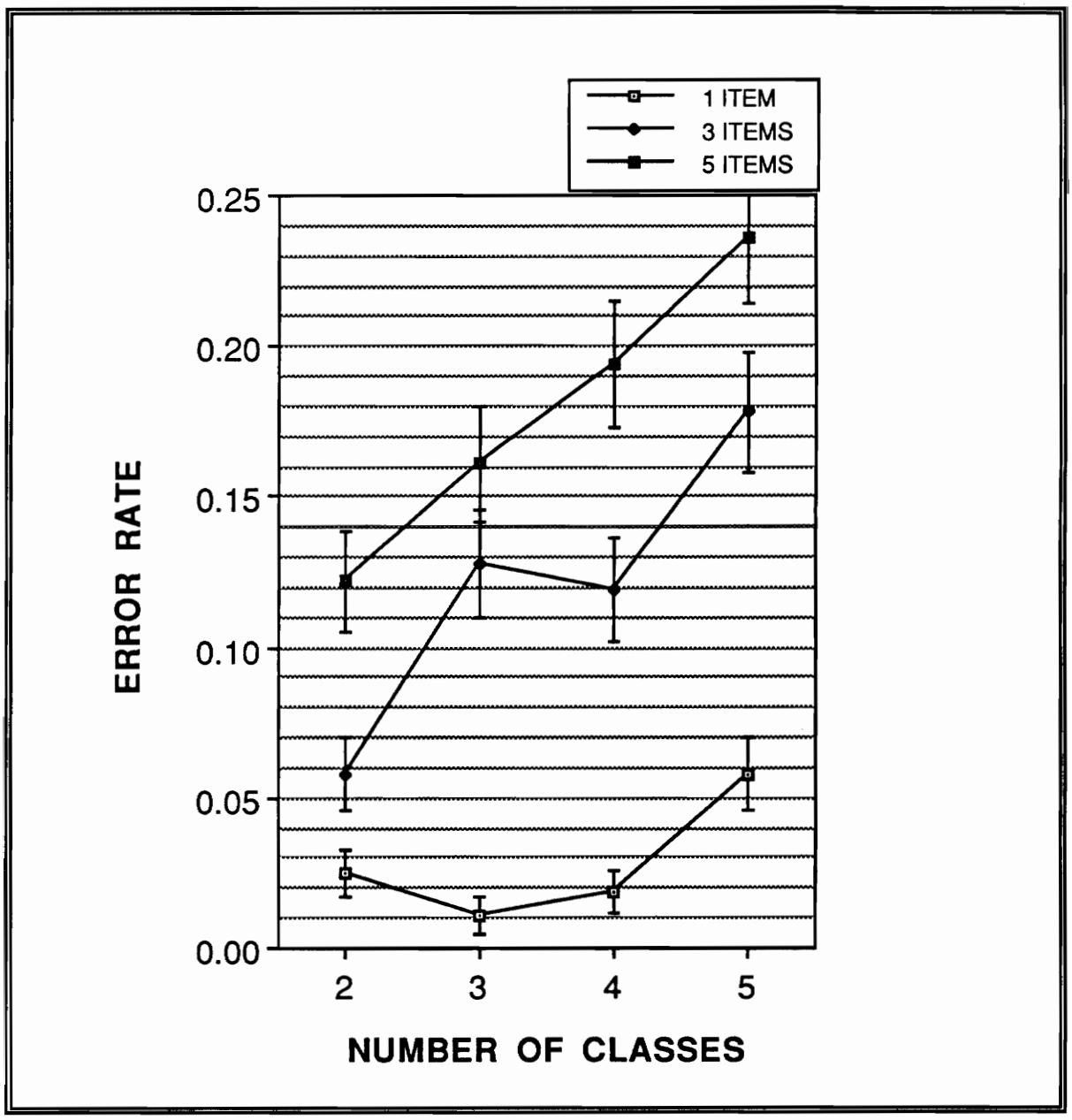


Figure 57. Effect of Number Of Classes X Number Of Items, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on error rates in Experiment 2 ($p < 0.0268$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

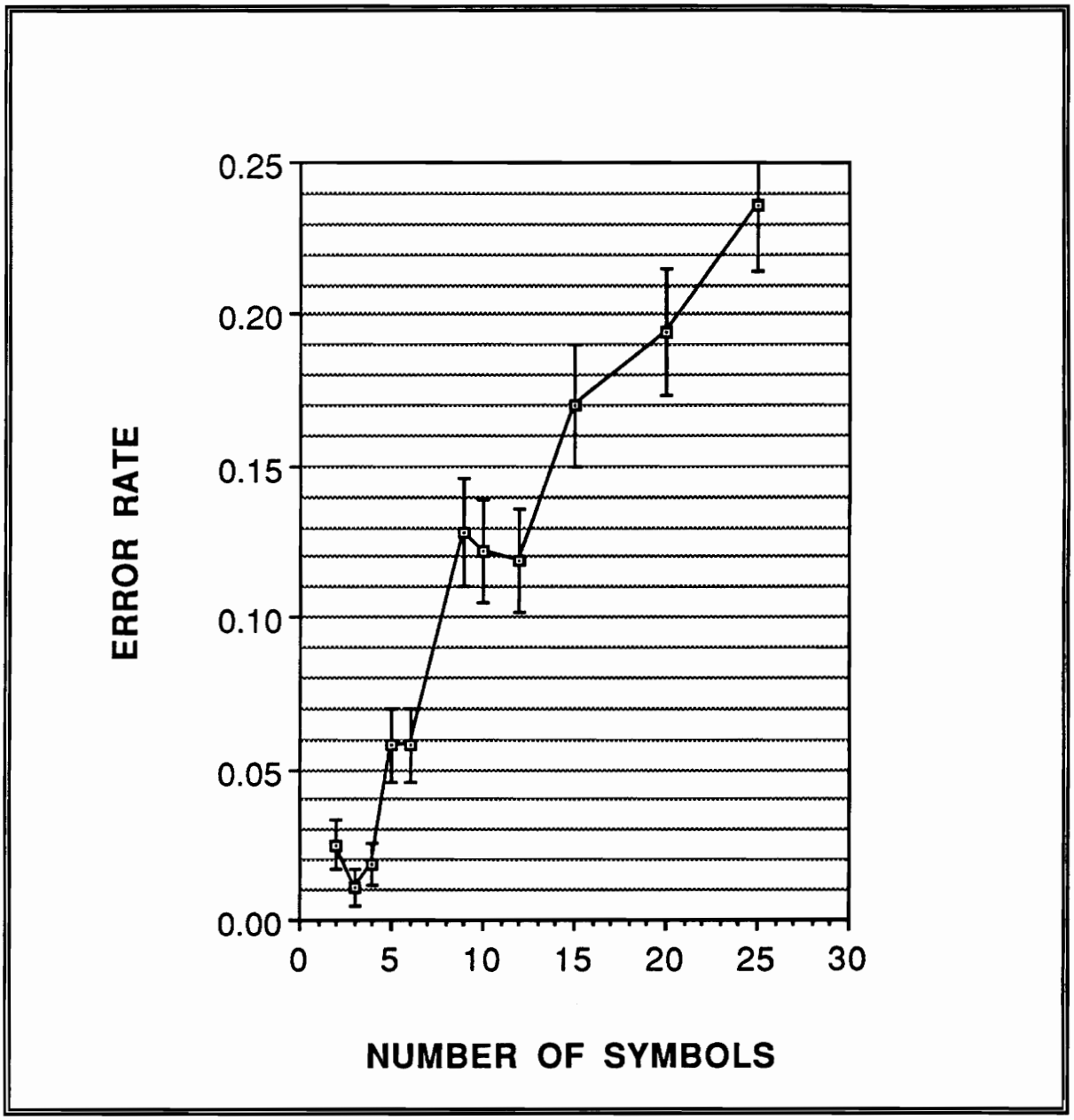


Figure 58. Effect of number of symbols, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on error rates in Experiment 2. Error bars indicate $\pm$ 1 $s_{\bar{x}}$. Note: There were two cases in which the number of symbols equalled 15. Average means and $s_{\bar{x}}$s are shown.

levels of Spatial Proximity (p = 0.6599). Average error rates for the random, medium, and tight levels of Proximity are 0.112, 0.114, and 0.101, respectively.

It is interesting that treatments which grouped symbols provided significant and relatively large reductions in subjective workload, while providing no reduction in error rate. It is suggested that display grouping formats helped participants organize information into chunks, thus reducing subjective workload, yet participants were unable to effectively parse these chunks into their component items; thus, error rates remained constant. This view is supported by two lines of reasoning. First, debriefing interviews indicated that participants always attempted to chunk displayed information. Second, results from Experiment 1 indicate that grouping formats produce lower error rates and subjective workload for an item categorization task similar to that required as part of the Experiment 2 task.

Although previous research has found that chunking reduced error rates, it is suggested that there are two kinds of chunks: STM-LTM chunks and STM chunks. STM-LTM chunking indicates use of a smaller number of labels stored in STM to access a larger number of items stored in LTM. STM-LTM chunks were described earlier in this dissertation (Figure 1). For example, Smith (1954; cited by Miller, 1956) employed a 4:1 recoding scheme such that the binary digits 0, 1, 1, 1 were stored in STM as the digit "7". After recalling the chunk label "7" from STM, Smith applied the recoding scheme to regenerate the original binary digits. Similarly, Chase and Simon's (1973) expert used chunk labels such as "castle formation," where each label held the identity and location of numerous individual chess pieces.

As discussed earlier, STM-LTM chunks may be available only to experts in a particular field. It has been suggested that an expert is an individual who maintains a library of learned information patterns and appropriate responses to these patterns. Therefore,

expertise allows increased memory of presented situations. Two additional advantages of expertise (i.e., STM-LTM chunking) are perhaps even more important. First, the expert is able to perceive incoming information in terms of known chunks, and, hence, information can be organized more quickly. Second, experts understand the appropriate responses to a particular chunk. Therefore, an expert does not need to calculate a response "on the fly".

The second type of chunking, STM chunking, was used by Experiment 2 participants. Here, individuals organized related items in STM, but they did not have well-learned LTM patterns with shorthand labels, such that each STM label could be used to prompt recall of a specific set of LTM items. Therefore, participants organized incoming information into chunks of related items, but all of the to-be-remembered items were stored in STM (i.e., none were "off-loaded" to LTM). For example, the CIS Order might have helped a participant encode and store the memory string *bear, dog, fork, spoon*. This string indicates two chunks of information: "animals" and "kitchen items". In the Parallel Order conditions, participants may have encoded and stored this same string as *bear, fork, dog, spoon*. This second string indicates four chunks of information, with each chunk containing a single item. Regardless of the number of organizing chunks, each of the to-be-recalled items was held in STM and occupied one STM slot.

It was noted earlier that some participants attempted to create shorthand labels. For example, the symbols *cup* and *pitcher* might have been encoded as "handle". Given more practice, participants would have been able to create such chunks and use them effectively. However, the chunk labels available to them (e.g., "kitchen") did not provide one-to-one correspondence between labels and items. For example, the label "kitchen" might indicate *cup* and *pitcher* or *fork* and *spoon*. Still, these chunks would have reduced error rates if negative response probes had been selected from classes which had not been presented.

Experiment 2 debriefings indicated that STM chunks, like STM-LTM chunks, are organized according to relations stored in LTM. For example, participants preferred the chunk *1, 2, 3, 4* to the chunk *3, 1, 4, 2*, presumably because the former organization was consistent with a well-known LTM pattern. It seems plausible that the frequency of LTM pattern usage leads individuals to assign shorthand nomenclatures. If participants repeatedly encounter the numerals 1 through 4, they would create a shorthand label for this pattern. Therefore, it is suggested that an STM chunk is an STM-LTM chunk in its infancy. The operator has imposed an order, but not yet provided a shorthand label.

In summary, it is suggested that participants sought to organize incoming information as a step towards developing expertise at the Experiment 2 task. Treatments which facilitated grouping reduced subjective workload because they helped participants organize information. Display grouping formats did not reduce error rates because participants did not have time to develop well-learned, quickly applicable STM-LTM chunks. Finally, it should be noted that rapid chunk formation (as was demonstrated in the Experiment 1), not recall of chunk components from memory, may be of greater value to operator tasks. Second, while chunking did not improve detailed component recall, it also did not degrade component recall. Third, given better-trained participants, longer task viewing times, or a task in which negative probe items were selected from non-displayed classes, display grouping may have reduced error rates as well as subjective workload.

Combined effects of complexity and grouping. There are no significant error rate results suggesting that display grouping effects varied with display complexity.

Effects due to Type. There are no significant results suggesting that Type effected error rate.

Response Time

The average response time across all conditions examined was 1269.22 ms.

Effects due to display complexity. As illustrated in Figure 59, response time increases monotonically with increasing Number Of Classes ($p < 0.0001$). A Newman-Keuls test (Table H 19) indicates that some, but not all, means differ from one another. In Figure 60, response time increases with increasing Number Of Items ($p < 0.0001$). Again, a Newman-Keuls test (Table H20) indicates that some, but not all, means differ from one another.

In Figure 61, response time is plotted against the total number of symbols. It is apparent that when the Number Of Items is six or fewer, response time increases more rapidly with increasing numbers of symbols. When the number of symbols is greater than six, response time increases more gradually. It is suggested that when the total number of symbols is small, participants created an MSET string of the items and, after presentation of the probe, serially scanned this MSET string. As the number of symbols increased, participants created and scanned an MSET string containing a subset of all displayed items. Therefore, the response time (per symbol) slope declined. This is consistent with Miller's (1966) belief that people can store approximately seven chunks of information.

Effects due to display grouping. There are no significant main effects of the display grouping variables upon response times.

Combined effects of complexity and grouping. Four interactions indicate that the effects of display grouping on response time vary with display complexity. These interactions are: Number Of Classes X Presentation Order ($p < 0.0773$), Number Of Items X Presentation Order ($p < 0.0544$), Number Of Classes X Number Of Items X Presentation Order ($p < 0.0220$), and Number Of Classes X Number Of Items X Presentation Order X Proximity ($p < 0.0592$) (Figures 62 through 64).

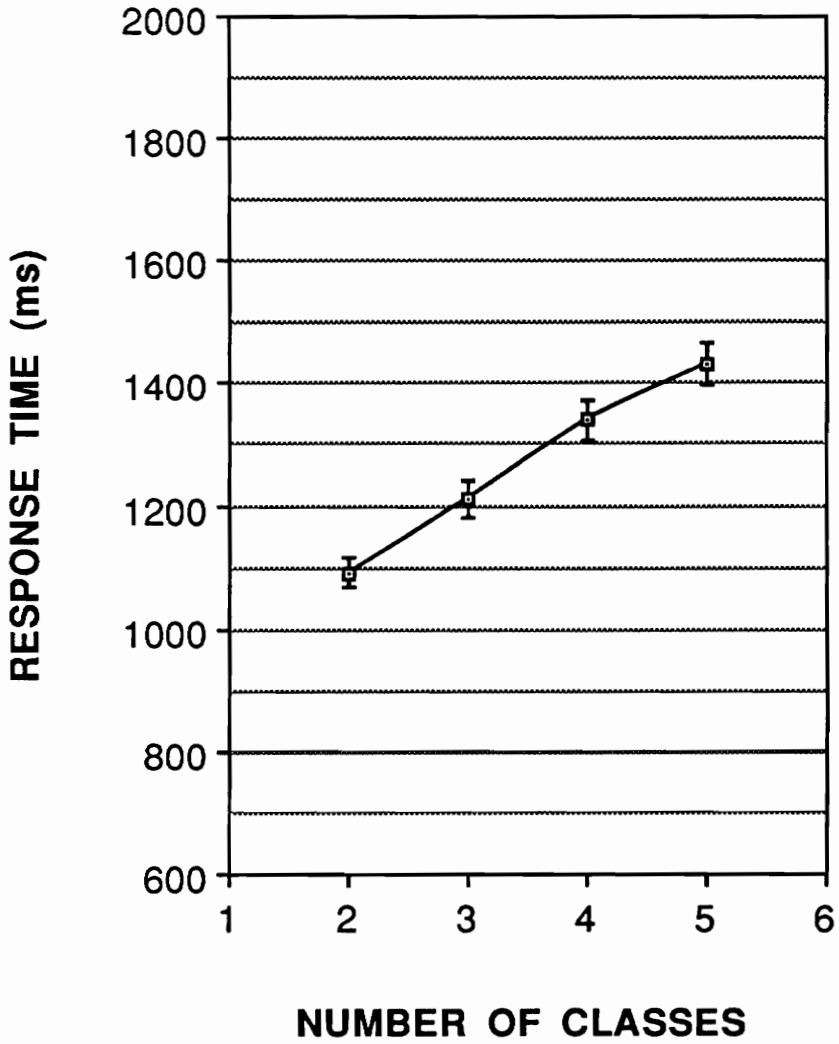


Figure 59. Effect of Number Of Classes, collapsed over subjects (N = 10), Blocks, Number Of Items, Presentation Order, Proximity, and Type, on response times in Experiment 2 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

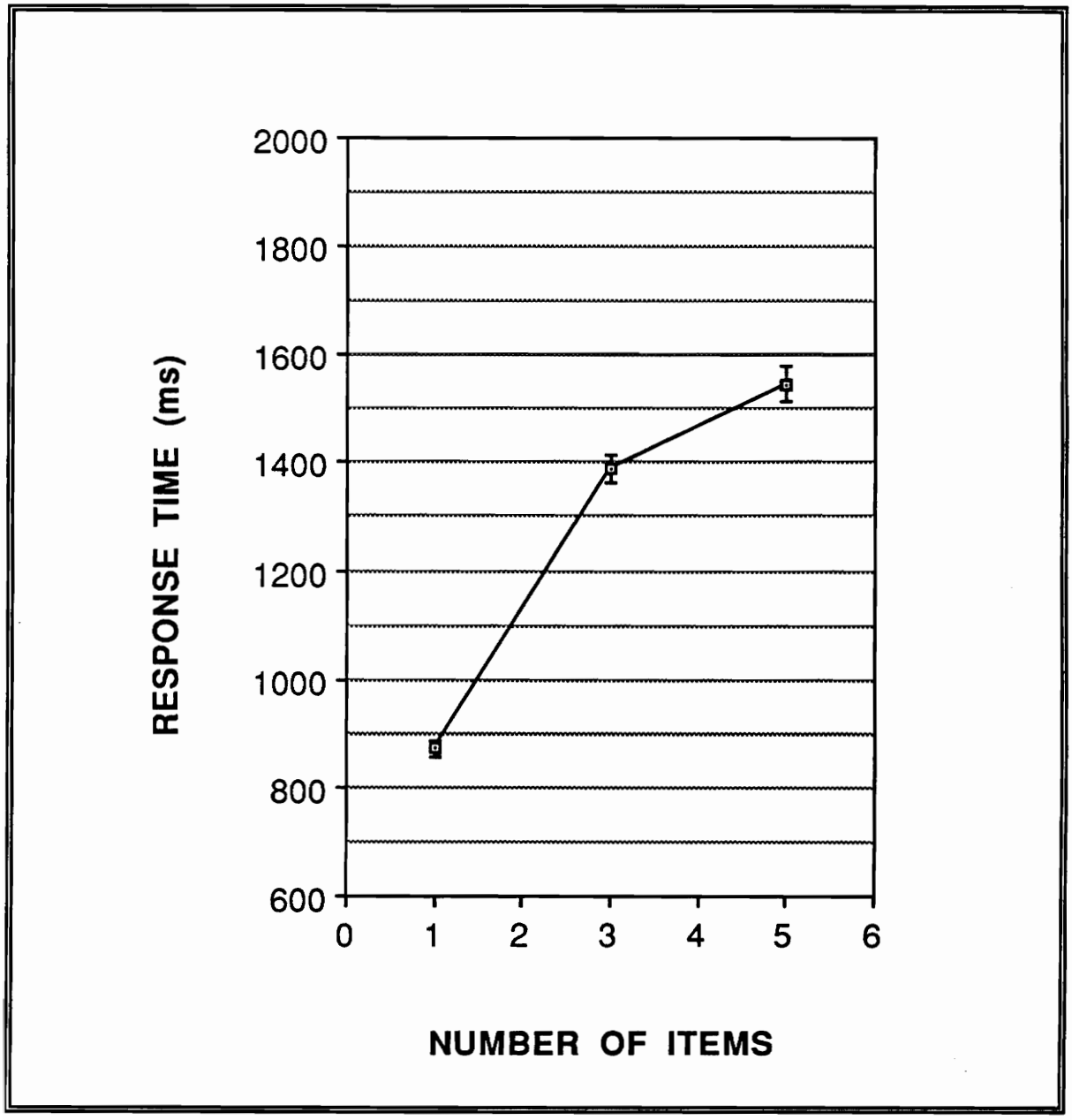


Figure 60. Effect of Number Of Items, collapsed over subjects (N = 10), Blocks, Number Of Classes, Presentation Order, Proximity, and Type, on response times in Experiment 2 ($p < 0.0001$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

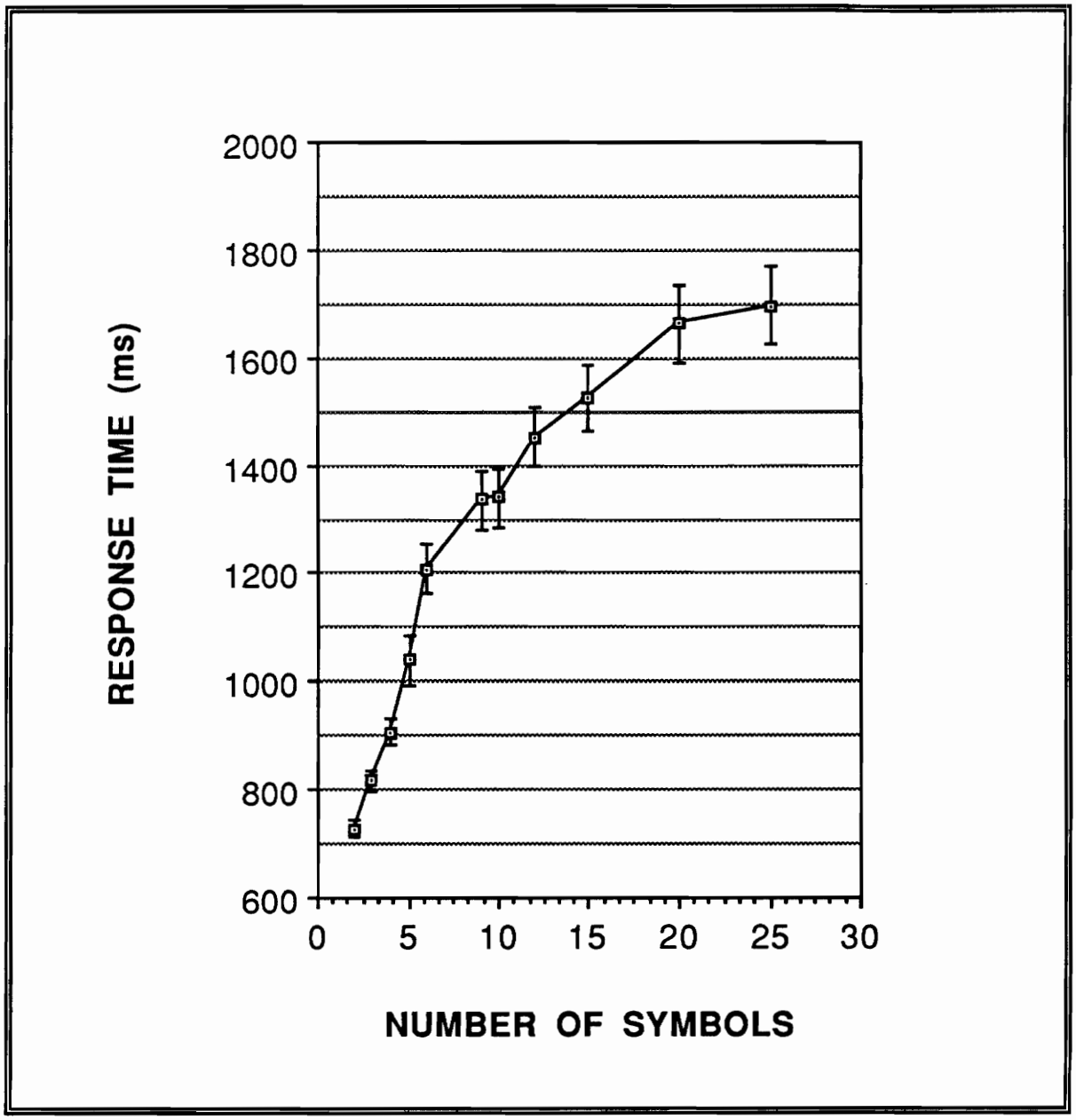


Figure 61. The effect of the number of symbols, collapsed over subjects (N = 10), Blocks, Presentation Order, Proximity, and Type, on response times in Experiment 2. Error bars indicate $\pm 1\ s_{\bar{x}}$. Note: There were two cases in which the number of symbols equalled 15. Average means and $s_{\bar{x}}$ s are shown.

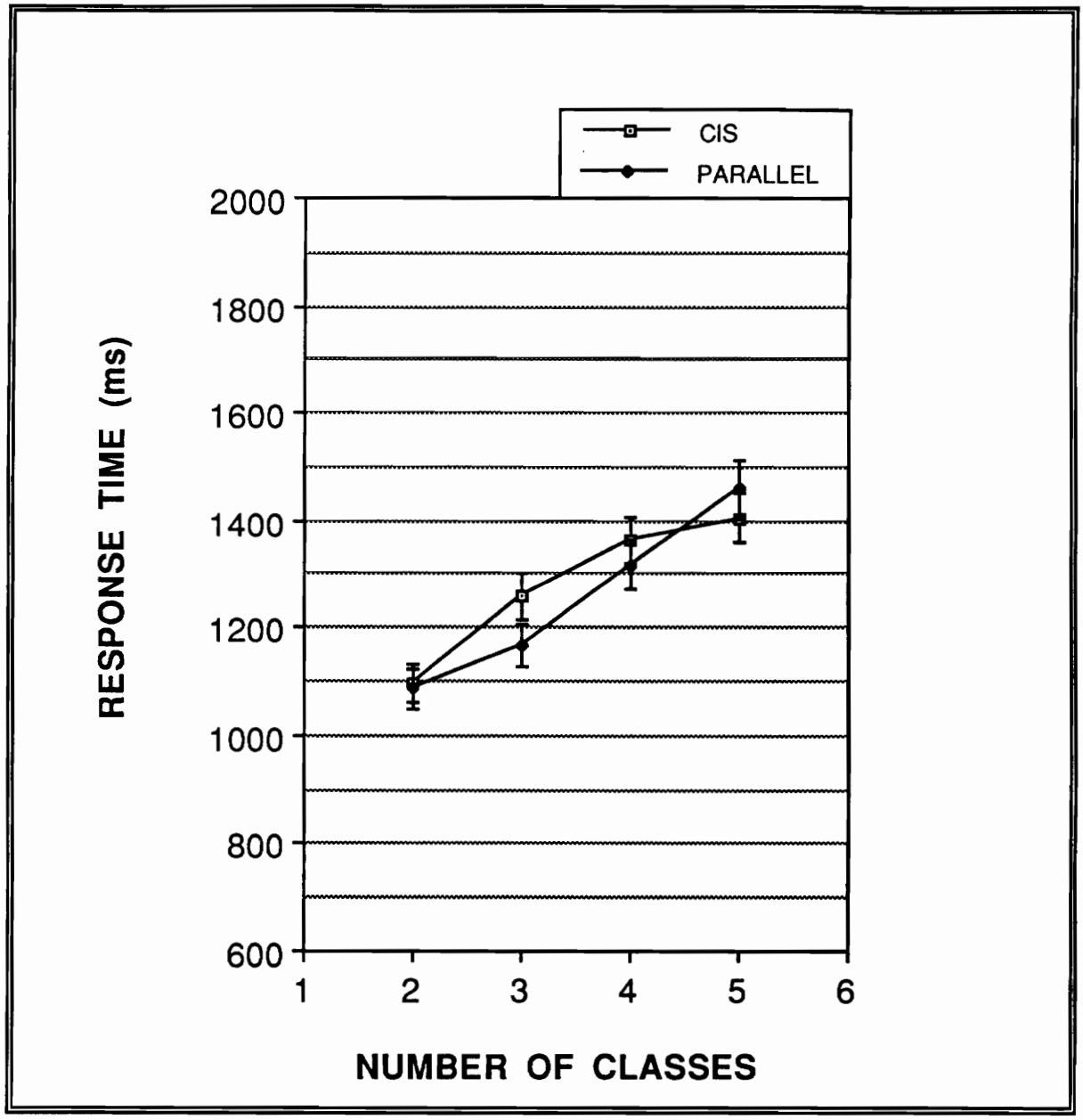


Figure 62. Two-factor interaction of Number Of Classes X Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Items, Proximity, and Type, on response times in Experiment 2 ($p < 0.0773$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

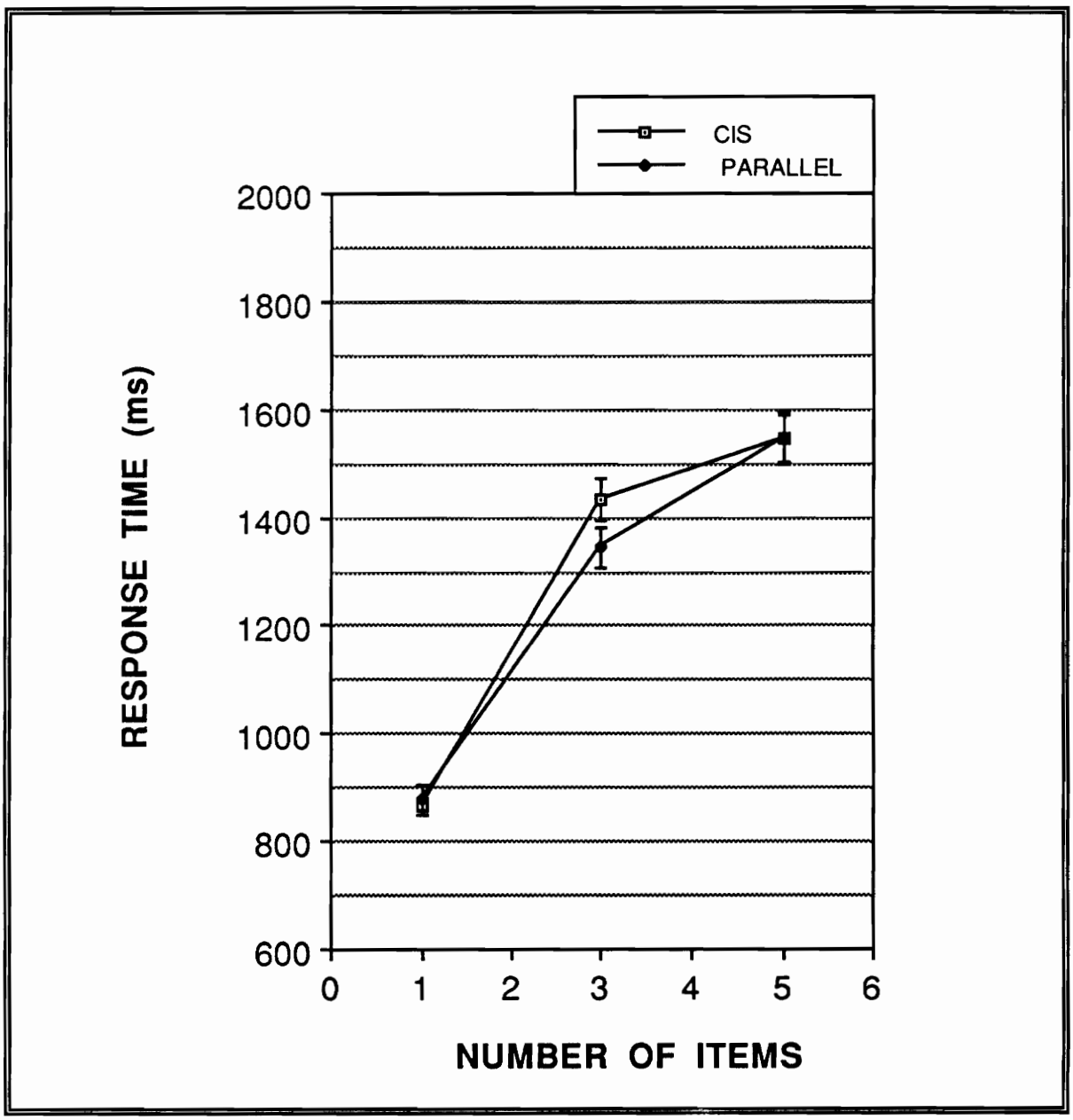


Figure 63. Two-factor interaction of Number Of Items X Presentation Order, collapsed over subjects (N = 10), Blocks, Number Of Classes, Proximity, and Type, on response times in Experiment 2 ($p < 0.0544$). Error bars indicate $\pm 1\ s_{\bar{x}}$.

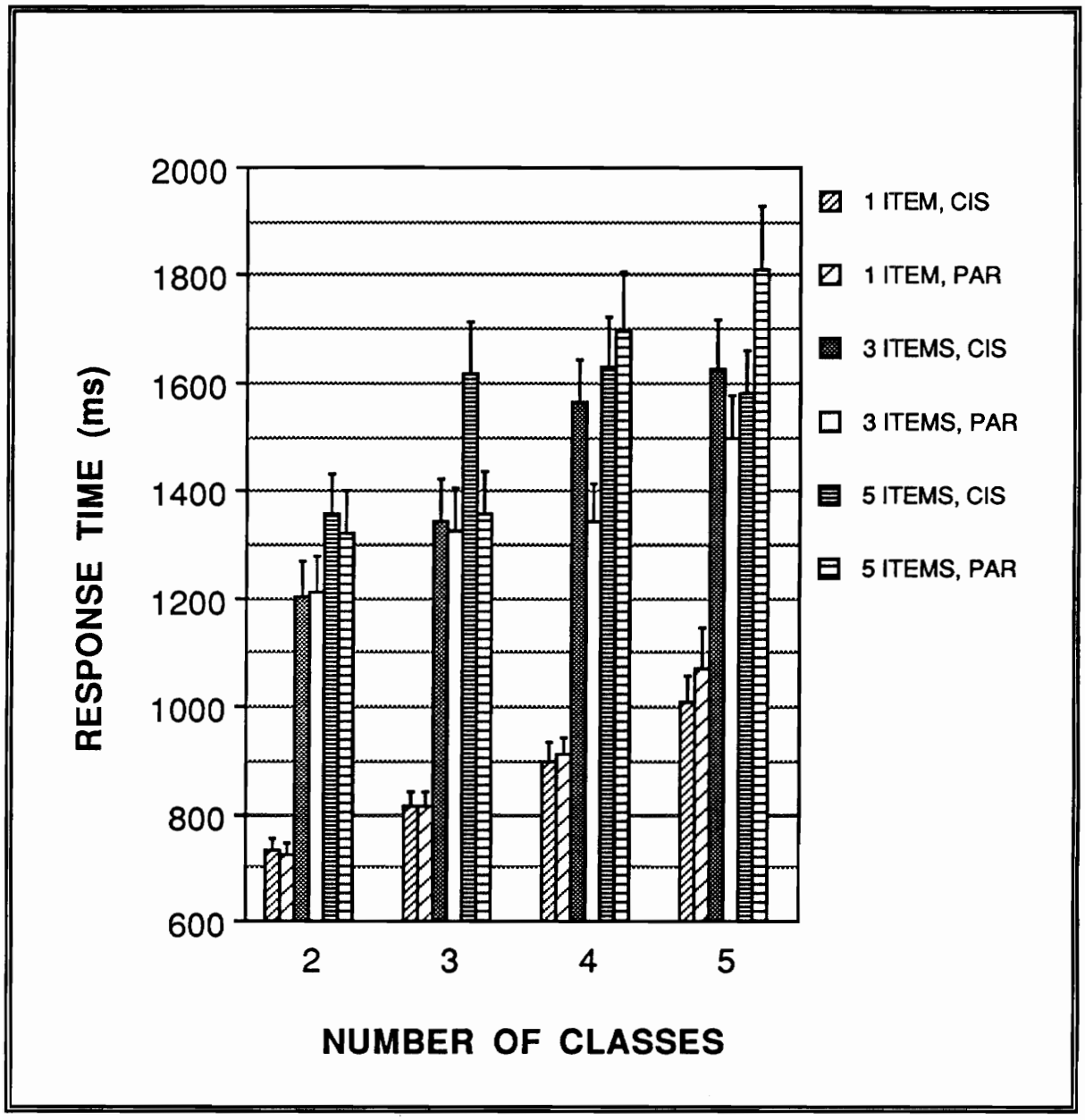


Figure 64. Three-factor interaction of Number Of Classes X Number Of Items X Presentation Order, collapsed over subjects (N = 10), Blocks, Proximity, and Type, on response times in Experiment 2 ($p < 0.0220$). Error bars indicate + 1 $s_{\bar{x}}$.

In general, the data trends observed in these interaction effects do not alter findings based on component (i.e., lower order) effects. Findings that response time increases with display complexity have not been modified. Similarly, no display grouping main effects or practically useful trends are indicated.

Effects due to Type. Finally, the main effect of Type is significant ($p < 0.0110$). As illustrated in Figure 65, the average response time is 1338.24 ms for negative trials and 1200.20 ms for positive trials. This Type effect is consistent with earlier results. Again, it is suggested that with positive trials, a participant could match the probe to an MSET item and terminate the MSET search. Upon failing to make a positive identification, participants scanned exhaustively all MSET items and initiated an educated guess procedure in order to select a response.

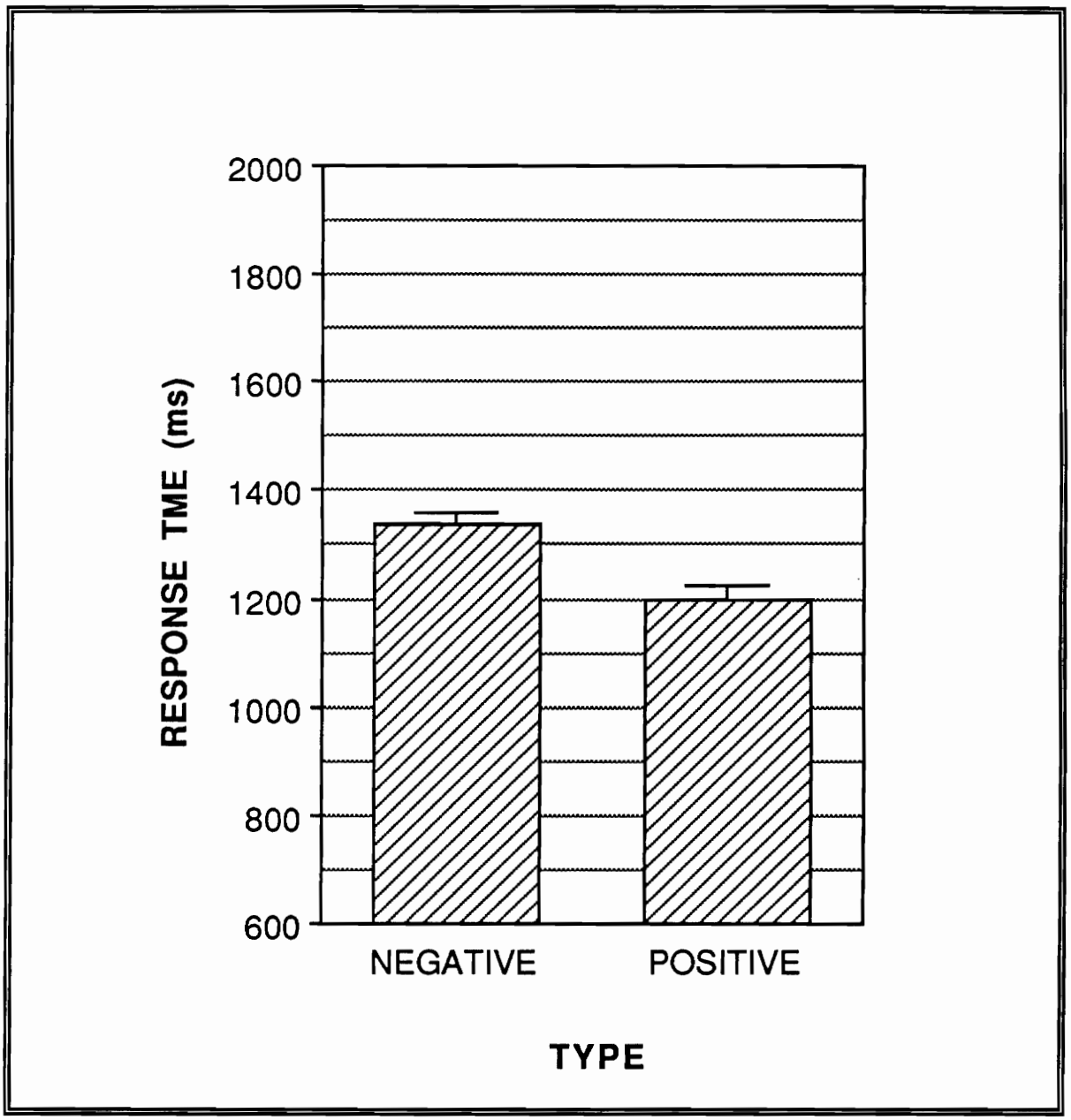


Figure 65. Effect of Type collapsed over subjects (N = 10), Blocks, Number Of Classes, Number Of Items, Presentation Order, and Proximity, on response times in Experiment 2 ($p < 0.0220$). Error bars indicate + 1 $s_{\bar{x}}$.

CONCLUSIONS

In general, Experiment 1 indicated that display grouping formats can help operators rapidly categorize information into appropriate cognitive chunks. Display grouping formats reduced both error rates and subjective workload. Experiment 2 indicated that grouping formats can reduce subjective workload required to categorize, store, and parse these chunks, but may not always reduce error rates. Thus, it has been demonstrated that grouping formats can be a valuable aid to reducing operator workload associated with complex information displays and that this line of research should be continued.

Summary

This research has studied the method by which operators organize displayed information in order to make decisions. It has been suggested that experts, unlike novices, have existing patterns of task-relevant information stored in LTM. Incoming information is compared against these patterns and pre-learned responses are implemented. It has also been suggested that electronic displays aided by artificial intelligence are able to manipulate the way information is presented and therefore could present information in a matter that allows appropriate and efficient recognition of relevant information chunks.

The most important contribution of this research has been the introduction and assessment of two grouping methods which might be incorporated into the design of electronic displays. The first, Chunks-In-Sequence (CIS), consists of grouping displayed information in the temporal dimension. That is, a set of conceptually related items (i.e., a chunk) is presented and blanked out; then, a second set is presented and blanked out, etc. The second method, Tight Spatial Proximity, consists of altering the physical location of displayed items so that conceptually related items are located close to one another.

Research Findings

Two experiments were conducted. Experiment 1 assessed whether display grouping formats (i.e., CIS and Tight Proximity) could help participants categorize information into appropriate chunks. The research hypothesis was supported in that grouping formats enabled participants to complete the task with lower subjective workload ratings and error rates.

Experiment 2 assessed the effect of the display grouping formats upon participants' abilities to create chunks, store the chunks, and then parse the chunks into their component items. The research hypothesis was partially supported. That is, grouping formats allowed participants to complete task operations with lower subjective workload, but with equivalent error rates to non-grouped formats. It was suggested that grouping formats helped participants categorize information into chunks, thus reducing subjective workload. Yet, grouping did not help participants parse these memorized chunks into component items.

Finally, it was suggested that there are two types of chunks. STM-LTM chunking indicates use of a smaller number of labels stored in STM to access a larger number of items stored in LTM. This type of coding is available to experts at a given task. STM chunking consists of organizing related items in STM. Here, individuals do not have access to extant LTM patterns. Therefore, they can not off-load to-be-remembered items to LTM and STM span is not increased.

Regarding Experiment 2, it is believed that display grouping would have produced lower error rates given slightly different task parameters. First, participants did not have adequate training or practice time to become expert at the task. Given training and/or practice they would have had extant STM-LTM chunks (e.g., *1, 3, 5* equals "odd") and

been practiced in using them. These STM-LTM chunks would have allowed storage of more items, given constant STM spans. Second, the Experiment 2 task allowed brief viewing exposures. Participants might have created STM-LTM chunks given longer viewing times. Third, the Experiment 2 task required memorization, not problem solving. It is likely that STM chunks allow enhanced real-time problem-solving performance (e.g., consider Experiment 1 results) while not enhancing short-term memorization performance.

Design Implications

This research clearly indicates that grouping formats can help operators rapidly categorize information into appropriate cognitive chunks. However, it is equally apparent that there are issues which must be addressed before grouping designs are incorporated.

Chunking versus parsing chunks. Experiment 1 suggested that display grouping greatly enhances categorization performance, while Experiment 2 suggested that grouping formats may not aid parsing of elements held in these chunks. It is likely that rapid recognition of chunks (whether STM-LTM or STM) and implementation of known responses is paramount to operator performance. The ability to parse items from memorized chunks will be called upon less often. First, operators will most often be able to view individual symbols on their displays. Second, the premise of this research effort is that operators must respond to patterns of information, not individual items.

Chunkable versus non-chunkable information. It has been suggested that when there is only one item of information per appropriate (i.e., task-relevant) chunk, display grouping will not provide a benefit. Conversely, as the number of items per appropriate chunk increases, benefits associated with display grouping and cognitive chunking will also increase.

CIS Versus Positional Chunking. It is tempting to ask: *which is better, CIS or Tight Proximity?* Across the experimental tasks, the two display grouping methods produced nearly identical results. Yet, the experiment design process and participants' debriefing comments suggest that, given alternative task situations, CIS or Tight Proximity might prove superior.

With extremely brief information display, CIS presentation may be undesirable for a number of reasons. In Experiments 1 and 2, chunks of information appeared and disappeared and this was controlled by the computer (externally paced relative to the participant). Some participants commented that each chunk (i.e., screen) was displayed only briefly. Therefore, if they got a slow start, perhaps due to a mental lapse, it was difficult to recover. Some participants also commented that they sometimes felt they were allowed too much time to process one chunk and not enough for another. (One possible solution would allow operators to manually override CIS presentations; however, additional system complexity would be introduced.)

Finally, it is possible that CIS presentation might allow operators enhanced understanding of individual chunks, while inhibiting understanding of relations among chunks. Thus, CIS presentation might degrade performance of an operator who is required to recognize and respond to multi-chunk patterns. As Experiments 1 and 2 required participants to make decisions based upon individual chunks, the affect of display grouping strategies upon multi-chunk pattern recognition was not assessed.

Perhaps the greatest advantage of CIS, relative to Tight Proximity, grouping is the maintenance of spatial verity. In some instances it may be necessary to present items in exact physical locations.

Positive versus negative searches. Participants' performance (i.e., error rate, subjective rating, response time) was slightly better for trials in which they could respond that a

particular item had been displayed. Performance was lower when they reported that an item was not present. This trend suggests a possible advantage to displaying positive case data, but also positively displaying negative case data (i.e, display the set of items which was not presented on the primary display). Advantages of displaying negative case data do not necessarily outweigh potential display clutter disadvantages.

Research Issues

This research employed two different tasks and indicated benefits associated with display grouping formats. Experiments which vary task parameters should be conducted. This additional study will allow researchers to assess the robustness of the grouping/chunking phenomenon. For example, Experiments 1 and 2 employed a Sternberg paradigm which, in essence, required participants to organize and memorize displayed items. Future research should use tasks which require participants to view displayed information, solve problems, and input responses, all in real-time. Some suggested generic tasks are provided in Table 1. As suggested above, one such task would require recognition of patterns across multi-chunk displays.

Second, Experiments 1 and 2 used novice participants. It is important to test participants who have had training and/or experience in performing the experimental task. Experts will have extant chunks which may affect the utility of display grouping methods. Third, it has been noted that items should be grouped in a manner appropriate to the task at hand. It is important to understand performance factors which relate to inappropriate display grouping.

Finally, it has been suggested that operators should be trained to recognize and respond to chunks, not to individual symbols. It would be interesting to take an existing training program and retool it in terms of the philosophy that *expertise is the ability to rapidly*

recognize and implement responses to extant chunks. Performance of operators who received alternative training methods could then be tested.

REFERENCES

Bartram, D. J. (1980). Comprehending spatial information: The relative efficiency of different methods of presenting information about bus routes. *Journal of Applied Psychology*, *65*(1), 103-110.

Cavanagh, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, *79*(6), 525-530.

Charness, N. (1985). Aging and problem solving performance. In N. Charness (Ed.), *Aging and human performance* (pp. 225-229). Chichester: Wiley.

Chase, W. G., and Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.

de Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving: Research method and theory.* (pp. 19-50). New York: Wiley.

Eisen, P. S. and Hendy, K. C. (1987). A preliminary examination of mental workload, its measurement and prediction. (DCIEM Report No. 87-RR-57). Downsview, Ontario: Defence and Civil Institute of Environmental Medicine.

Ellis, N. C. and Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, *71*, 43-51.

Goodstein, L. P. (1981) Discriminative display support for process operators. In J. Rasmussen and W. B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 433-449). New York: Plenum Press.

Krantz, D. H., and Tversky, A. (1971). Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, *78*, 151-169.

Kuperman, G. G., and Wilson, D. L. (1986). *A human-centered technology for advanced bomber crew stations* (Report No. AAMRL-TR-86-052). Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Medin, D. L., and Smith, E. E. (1984). Concepts and concept formation. *Annual review of psychology, 35,* 113-138.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Nygren, T. E. (1982). *Conjoint measurement and conjoint scaling: a users guide* (Report No. AFAMRL-TR-82-22). Wright-Patterson AFB, Ohio: U. S. Air Force Aerospace Medical Research Laboratories.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics, 13*(3), 257-266.

Reid, G. B., Potter, S. S., and Bressler, J. R. (1989, July) *Subjective Workload Assessment Technique (SWAT): A User's Guide (U).* (Report No. AAMRL-TR-89-023). Wright-Patterson AFB, OH: Armstrong Aerospace Medical research Laboratory.

Reinhart, W. F., Glynn, C. D., Dye, C., Takahama, M. S., and Snyder, H. L. (1988). *The role of short-term memory in operator workload.* Virginia Polytechnic Institute and State University, Blacksburg, VA: Department of Industrial Engineering and Operations Research.

Samet, M. G., Geiselman, R. E., and Landee B. M. (1980). *An experimental evaluation of tactical symbol-design features* (Report No. 498). Alexandria, VA: U. S. Army Institute for the Behavioral and Social Sciences.

Solso, R. L. (1988). *Cognitive psychology.* Boston: Allyn and Bacon.

Sternberg, S. (1966). High speed scanning in human memory. *Science*, *153*, 652-654.

Vistech Consultants. (1986).*VCTS application manual*. Dayton, OH.

Wickens, C. D. (1984). *Engineering psychology and human performance*. Columbus, OH: Charles E. Merrill.

Wierwille, W. W., and Connor, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. *Human Factors*, *25*, 1-16.

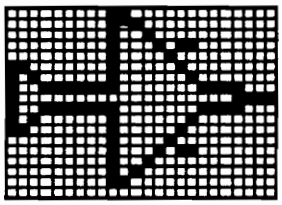
Williges, R. C., and Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. *Human Factors*, *21*, 549-574.

APPENDIX A

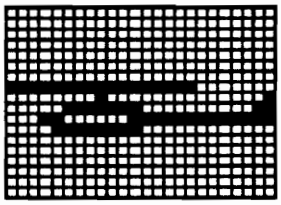
SYMBOL BIT MAPS

Weapon

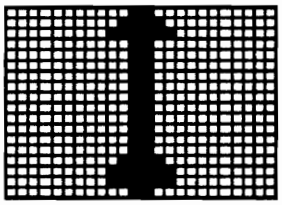
Airplane



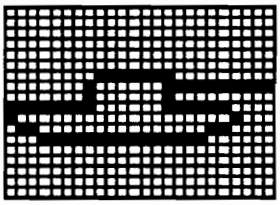
Helicopter



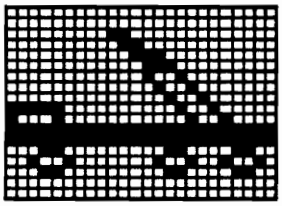
Missile (Air-To-Air)



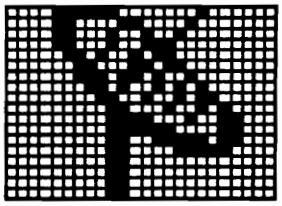
Tank



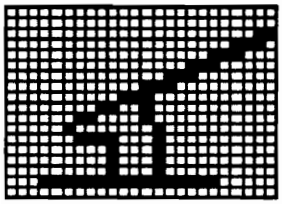
Truck (Truck Mounted Missile)



Radar



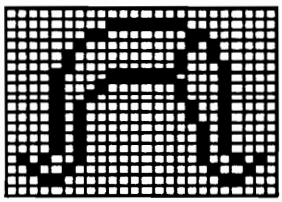
Gun



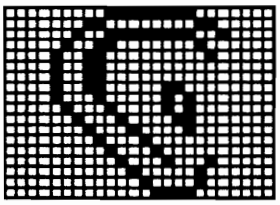
Warship

Human Body Part

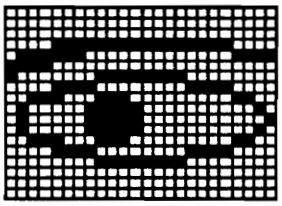
Hair



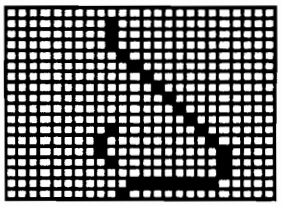
Ear



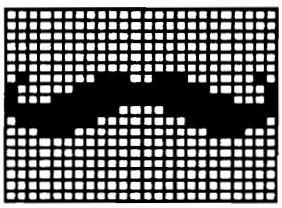
Eye



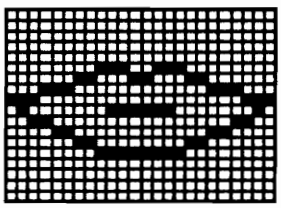
Nose



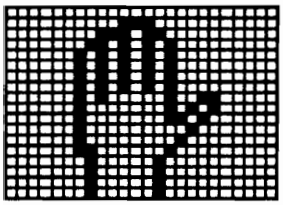
Mustache



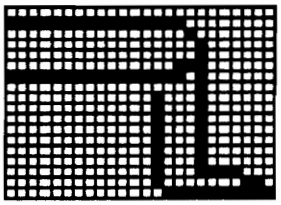
Lips



Hand

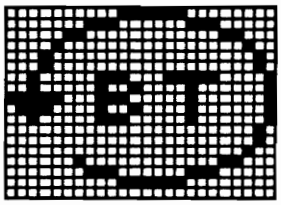


Leg

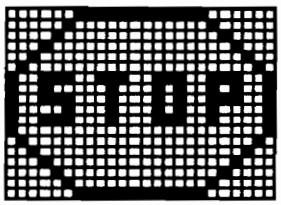


Sign (Traffic)

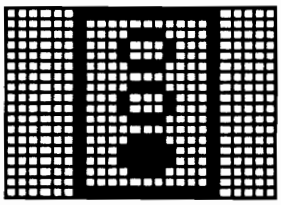
Blacksburg Transit
(Bus, BT)



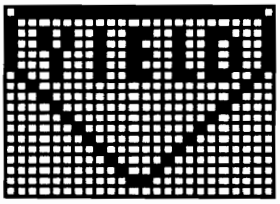
Stop Sign



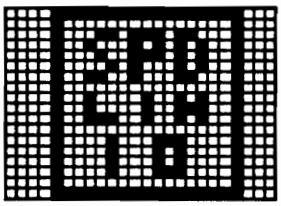
Stop Light



Yield Sign

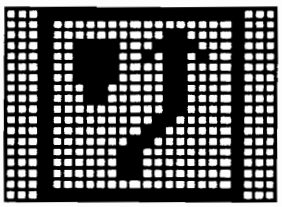


Speed Limit Sign

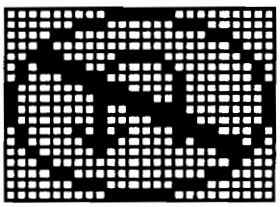


Highway Marker

Keep Right Sign

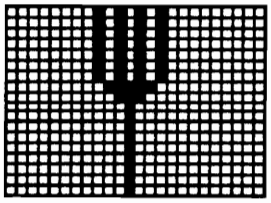


No-U-Turn Sign

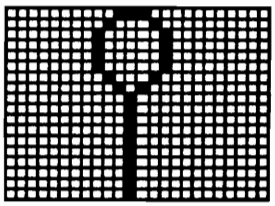


Kitchen

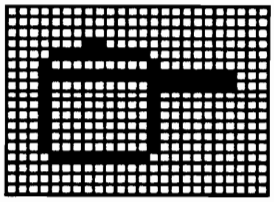
Fork



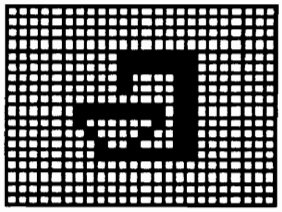
Spoon



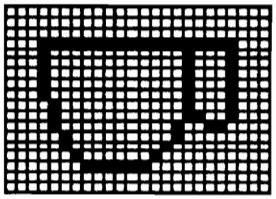
Pot



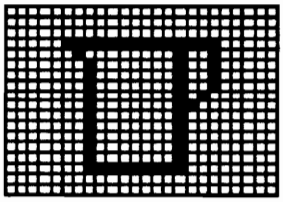
Mixer



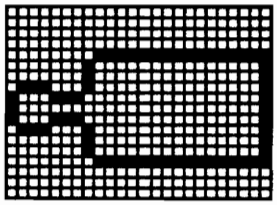
Cup



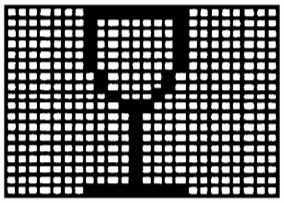
Pitcher



Breadboard

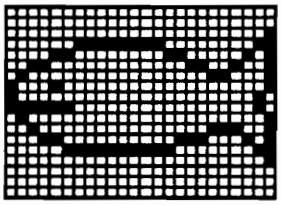


Goblet

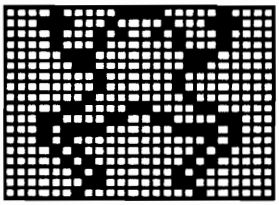


Animal

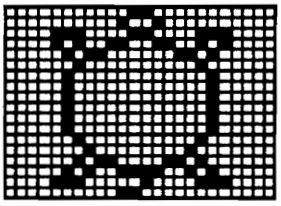
Fish



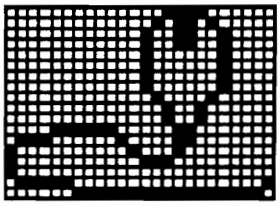
Frog



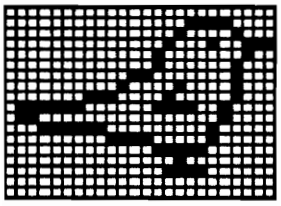
Turtle



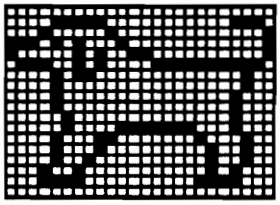
Snake (Cobra)



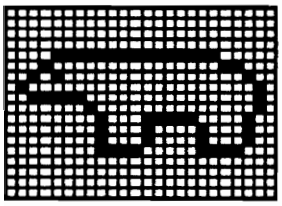
Bird



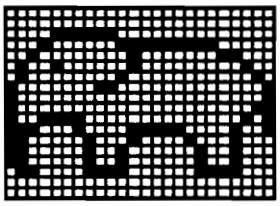
Dog



Bear

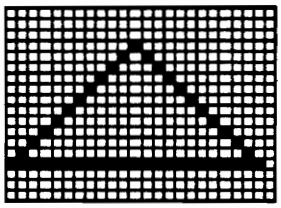


Elephant

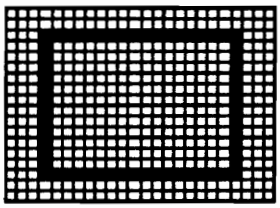


Shape (Simple)

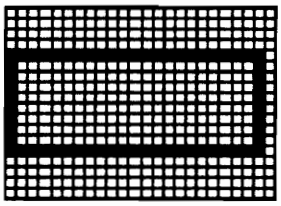
Triangle



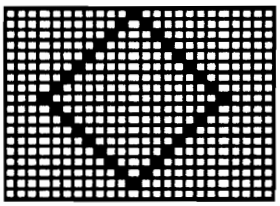
Square



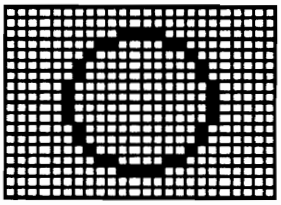
Rectangle



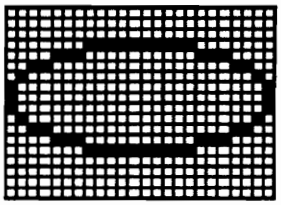
Diamond



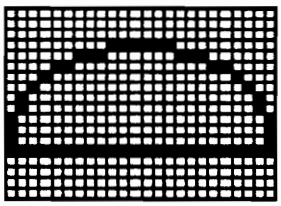
Circle



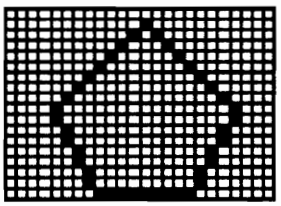
Oval



Semi-Circle

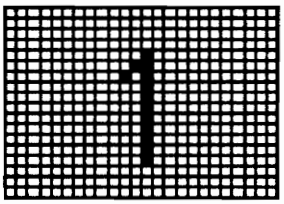


Pentagon

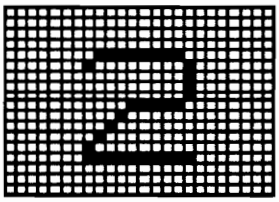


Number

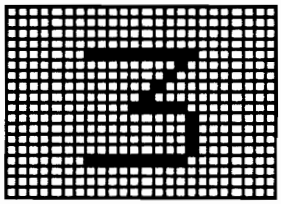
1



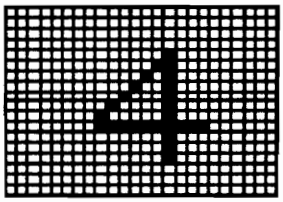
2



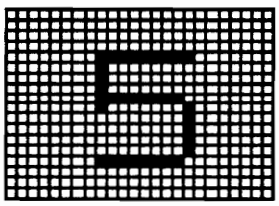
3



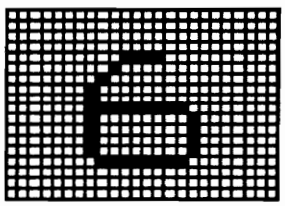
4



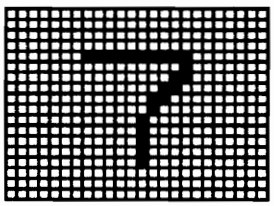
5



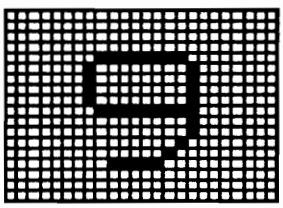
6



7

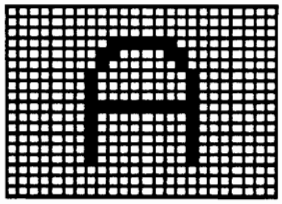


9

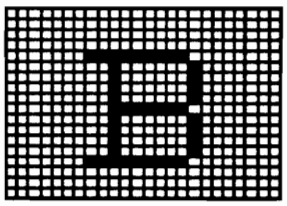


Letter

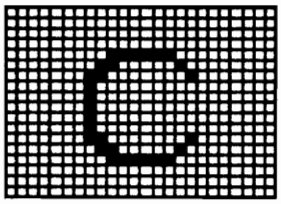
A



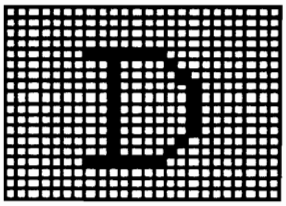
B



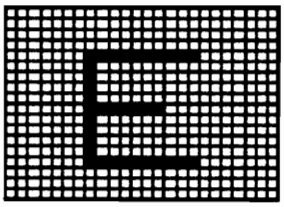
C



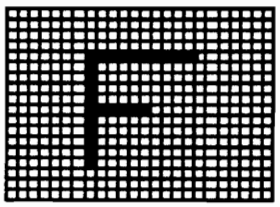
D



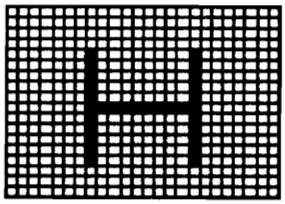
E



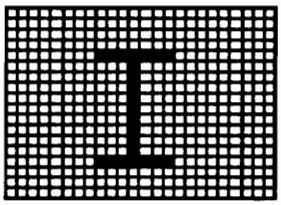
F



H



I



APPENDIX B

PARTICIPANT INFORMED CONSENT FORM

PARTICIPANT INFORMED CONSENT FORM

You are being asked to participate as a research participant in a research project. The purpose of this experiment is to examine performance given visual display tasks of different complexities. The task will require you to complete numerous trials. In each trial you will view a computer display and recall displayed information.

This research is being conducted by the Human Factors Laboratory of the Department of Industrial Engineering and Operations Research. IEOR research team members on this project include:

Kay C. Tan, Doctoral Student and Research Assistant (231-5499)
Michael J. Kahn, Doctoral Student and Research Assistant (231-5499)
Dr. Robert J. Beaton, Faculty Member and Principle Investigator (231-5936)

This study consists of two parts: screening tests and experiment sessions.

Screening Tests

First, you will complete a screening procedure. You will take a vision test. This test will require about 10 minutes in all. You will not be paid for completing the screening procedure.

If you are not invited to participate in the second portion of the experiment, you should not consider this as an inadequacy. We are seeking people with specific visual abilities. Also, as you are aware, people employ different perceptual styles to solve problems. We are seeking people who use a specific style. In fact, it is expected that many of the volunteers who apply will not be appropriate for this specific visual experiment.

Experiment Sessions

The second portion of the study is expected to last 3 days (starting today). The first session will last approximately 2.5 hours; subsequent sessions will be about 1.5 hour each. If you decide to participate, you will be paid $25.00 upon completing all 3 sessions. If you do not complete all 3 sessions, you will not be paid. This is because your data will be of no use to us.

During these sessions you will complete numerous trials. In each trial you will view a computer display and recall displayed information.
There are no known risks associated with this research. The only known discomfort to which you will be exposed is possible fatigue resulting from the length of the experiment. However, you will be permitted to take rest breaks when needed.

As a Research Participant, You Have Certain Rights:

1. It is your right to withdraw from the study at any time for any reason. However, as stated earlier, you will not be paid if you do not complete all 5 sessions.

2. One of the research team members will answer any questions you may have concerning this research. You should not sign this consent form until you are satisfied that you understand all the terms involved.

3. You have a right to see your data and withdraw it from the study if you so desire. If you decide to withdraw your data, please inform the experimenter immediately. Otherwise, identification of your data will not be possible since it is separated from the participant in order to ensure anonymity.

4. If you wish to receive a summary of the results of this research, please include your address (where you expect to be living three months from now) with your signature below. Please do so only if you are truly interested in seeing the results. If you desire more detailed information after receiving the synopsis, please contact the Human Factors Laboratory and a full report will be made available to you.

5. Should any further questions or problems arise, you may contact any of the research team members. If you have any concerns about the way the experiment is being conducted or the way you are being treated, you may contact Dr. E. R. Stout, Chairman of the Institutional Review Board (231-5281).

Your participation is greatly appreciated and we hope that you will find the study a pleasant and interesting experience. Your signature below indicates that you have read this document in its entirety, that your questions have been answered, and that you consent to participate in the study described. Further, your signature indicates that you will not

discuss participation in this study with anyone until June, 1990 when the study is to be completed.

_____

Please print name and address if you would like a synopsis of results

APPENDIX C

BRIEF EXPLANATION OF SWAT

BRIEF EXPLANATION OF SWAT

Mental workload describes a task's cognitive difficulty. By assessing the mental workload associated with alternative designs, the designer is better able to make design trade-off decisions. Mental workload metrics range from physiological (e.g., pupil diameter), to task analytic (e.g., time-line analysis comparing operator capabilities to task requirements), to behavioral. Behavioral metrics include primary task (e.g. error rates, response times), spare mental capacity, and subjective ratings. Subjective rating approaches require research participants to provide subjective opinions of task difficulty. Available methods include bipolar rating scales, magnitude estimation (i.e., participants assign numbers relating subjective difficulty), the Cooper-Harper scale, the NASA Task Load Index, and the Subjective Workload Assessment Technique (SWAT).

SWAT has dominated the area of subjective workload assessment (Eisen and Hendy, 1987) and has been applied across a wide variety of experimental situations. Per this method, subjective workload is composed of three dimensions: *time load*, *mental effort load*, and *stress load*. Each dimension is represented by a three-point rating scale.

SWAT is administered in two phases. The scale development phase is completed before participants are exposed to experimental treatments. Participants are given 27 index cards, each relating one of the possible combinations of *time*, *mental effort*, and *stress*. Each participant sorts these cards, to the order which he believes provides the easiest to the most difficult generic task. Conjoint analysis (Krantz and Tversky,1971; Nygren, 1982) is then performed upon participants' card sorts in order to convert these card sorts into a scaling solution(s). (Software which executes conjoint analysis is provided by Reid et al., 1987.) A scaling solution provides a combination rule that folds ratings on the three dimensions into a single indicant of workload. For example, the conversion rule might state that the

dimension ratings *time* = *1*, *mental effort* = *2*, and *stress* = *2* should be converted into the integer value 37.

The conjoint analysis software also checks participants' card sorts to ensure that each sort is internally consistent. Finally, this software provides a coefficient of concordance which indicates whether participants were homogeneous in their opinions as to the relative contributions of *time*, *effort*, and *stress* to the scaling solution. The experimenter must then determine whether to use a single overall scaling solution or multiple scaling solutions, with each solution being applied to a subset of the participants.

The second phase of SWAT application, event scoring, requires participants to rate the difficulty of specific tasks in terms of *time*, *effort* and *stress*. The previously developed scaling solution(s) is then applied to convert scores on these three dimensions into a single integer score.

Readers wishing to learn more about mental workload can refer to Eisen and Hendy (1987), Wierwille and Connor (1983), and Williges and Wierwille (1979). Readers wishing to learn more about the SWAT metric can refer to Appendix D and Reid et al. (1989).

APPENDIX D

SWAT CARD SORT INSTRUCTIONS TO PARTICIPANTS

SWAT CARD SORT INSTRUCTIONS FOR PARTICIPANTS

Performance Measures

We are interested in your performance on a number of tasks. Performance will be measured using task completion times and error rates. We also want your assessment of the amount of mental workload associated with each task.

Mental Workload

Mental Workload = A measure of mental effort. It is a measure of how hard you have to concentrate or work to complete a mental problem.

* For example, the amount of mental work required to solve a complex arithmetic problem is greater than that required to solve a simple one.
* We want to know how much and what kind of mental work you have to do to complete our tasks.
* One method of measuring mental workload is called SWAT (Subjective Workload Assessment Technique).

SWAT

SWAT describes mental workload in terms of three dimensions:

TIME LOAD

Amount of time pressure. Includes proportion of time you are busy, the degree to which tasks interfere with one another, and/or if there is not enough time to finish task(s).

* With a low Time Load, there is enough time to complete all mental work, plus some time to spare.
* With a high Time Load, you are very busy and task interruptions occur very frequently.

Time Load may be rated according to these three levels:

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.
3. Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.

MENTAL EFFORT LOAD

Amount of concentration required. Includes memorizing, performing calculations, or making difficult decisions.

* Regardless of the number of tasks that are performed at the same time.

* Regardless of any time limitations.

Mental Effort Load may be rated according to these three levels:

1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.

2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required.

3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

PSYCHOLOGICAL STRESS LOAD

How performance on a task is affected by factors such as anxiety, frustration, or confusion.

* At low levels, you feel relatively relaxed.

* As anxiety, frustration, or confusion increases, greater levels of concentration and determination are needed to perform the task.

Psychological Stress Load may be rated according to these three levels:

1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.

2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.

3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

CARD SORT PROCEDURE

Here is a stack of cards. Each card represents one level of mental workload combined from each of the three SWAT dimensions. For example, level 1 on Time Load, level 2 on Mental Effort Load, and level 2 on Psychological Stress Load combine to make one level of mental workload. Another level of mental workload would be represented by the combination: level 3 on Time Load, level 2 on Mental Effort Load, and level 3 on

Psychological Stress Load. Altogether, there are 27 cards representing the 27 combinations (i.e., 3 dimensions x 3 levels each) of the SWAT workload scale.

As part of the experiment, we need you to sort these 27 cards in the order in which *you* perceive as more or less mental workload. Sort in an increasing order from the lowest workload level to the highest workload level. Below is an example of a card sort and the resulting workload scale that will be obtained:

TIME	EFFORT	STRESS		WORKLOAD SCALE
1	1	1	---------->	0.0
.	.	.		.
.	.	.		.
.	.	.		.
3	3	3	---------->	100.0

POINTS TO REMEMBER WHEN DOING THE CARD SORT

1. When doing the card sort, use the descriptions printed on the cards. Use your general view of workload and consider the order of importance of each of the Time, Mental Effort, and Psychological Stress dimensions. Do not sort the cards based on only one task -- use your general view of mental workload.

2. During the experiment, you will be rating the mental workload involved in a series of trials. You will base you ratings on these three dimensions. For example, you may rate the mental workload involved in a particular trial as 1-2-2. This means 1 for Time, 2 for Mental Effort, and 2 for Psychological Stress.

3. We are not asking what you prefer. Some people prefer to be "busy" rather than "idle" in the Time dimension. Again, this is not what we are interested in. We want to know how the three dimensions and their three levels represent mental workload as *you* see it. You may prefer a 2-2-2 situation to a 1-1-1 situation which indicates that you prefer not to have too much spare time or that you work better under stress. Regardless, realize that 1-1-1 is **less** workload than 2-2-2.

4. **Please feel free to ask any questions as you are sorting the cards.**

APPENDIX E

PARTICIPANTS' INSTRUCTIONS FOR EXPERIMENT 1

PARTICIPANTS' INSTRUCTIONS FOR EXPERIMENT 1

In each experimental session you will perform a series of 144 memory trials. In each trial between 2 and 25 symbols will be presented.

Symbol Population.

Sixty-four symbols have been divided among eight symbol classes. Study these symbols so that you will be able to quickly identify each symbol and assign it to its correct class.

Your Task.

Each trial will require completion of the following steps:

1.) A symbol set will be displayed for a period of 10 seconds. You will view the displayed symbols and store (in memory) the class names (not the symbol names) of the symbols displayed. For example, if the screen shows the symbols "frog", "elephant", "airplane", and "helicopter"; you will remember "animal" and "weapon".

2.) After symbol presentation, a test item will be presented. The test item will be one of the 64 symbols (e.g., "turtle"). You will press the "yes" or the "no" key to indicate whether a symbol from that class was presented during that trial. Speed of response is important, but accuracy is more important. In the case where the symbols "frog", "elephant", "airplane", and "helicopter" were followed by the test item "bear", the correct response would be "yes". This is because "bear" belongs to the "animal" class.

3.) I will be changing the number of displayed symbols, their spatial arrangement, and the presentation timing (whether symbols are presented all a once, or sequentially). After each trial, you will rate how easy or difficult you thought it was to solve that problem given the display used. You should think about your rating carefully.

Practice.

Now, I will guide you through 10 practice trials to help you understand these instructions.

Remember.

Speed is important, but accuracy is more important.

APPENDIX F

TABLES FOR EXPERIMENT 1

TABLE F1

ANOVA Summary Table for SWAT Ratings (Experiment 1)

Source of Variance	df	MS	F	p
Number Of Classes (Nclass)	3	146894.81	41.60	0.0001
Nclass*Subjects	27	3531.53		
Number Of Items (Nitem)	2	71237.23	30.51	0.0001
Nitem*S	18	2335.05		
Presentation Order (Preso)	1	147631.84	26.13	0.0006
Preso*S	9	5649.69		
Proximity (Prox)	2	46078.66	26.61	0.0001
Prox*S	18	173.17		
Type	1	5694.59	8.06	0.0194
Type*S	9	706.21		
Nclass*Nitem	6	729.98	2.37	0.0421
Nclass*Nitem*S	54	167.55		
Nclass*Preso	3	111.10	0.38	0.7681
Nclass*Preso*S	27	292.24		
Nclass*Prox	6	540.67	1.87	0.1025
Nclass*Prox*S	54	288.74		
Nclass*Type	3	107.74	0.48	0.6966
Nclass*Type*S	27	222.85		
Nitem*Preso	2	35890.32	24.31	0.0001
Nitem*Preso*S	18	1476.66		
Nitem*Prox	4	13498.02	26.17	0.0001
Nitem*Prox*S	36	515.71		
Nitem*Type	2	1071.36	3.99	0.0369
Nitem*Type*S	18	268.81		
Preso*Prox	2	37353.50	25.72	0.0001
Preso*Prox*S	18	1452.14		
Preso*Type	1	216.86	0.94	0.3579
Preso*Type*S	9	231.01		
Prox*Type	2	631.93	1.49	0.2522
Prox*Type*S	18	424.56		
Nclass*Nitem*Preso	6	116.25	0.42	0.8606
Nclass*Nitem*Preso*S	54	274.89		
Nclass*Nitem*Prox	12	409.79	1.80	0.0574
Nclass*Nitem*Prox*S	108	228.00		
Nclass*Nitem*Type	6	155.05	0.57	0.7498
Nclass*Nitem*Type*S	54	270.44		
Nclass*Preso*Prox	6	137.82	0.57	0.7499
Nclass*Preso*Prox*S	54	240.48		
Nclass*Preso*Type	3	679.85	3.02	0.0470
Nclass*Preso*Type*S	27	225.01		

TABLE F1 (continued)

Nclass*Prox*Type	6	281.19	1.10	0.3717
Nclass*Prox*Type*S	54	254.52		
Nitem*Preso*Prox	4	7609.31	15.65	0.0001
Nitem*Preso*Prox*S	36	486.07		
Nitem*Preso*Type	2	144.13	0.55	0.5891
Nitem*Preso*Type*S	18	264.46		
Nitem*Prox*Type	4	278.84	1.24	0.3100
Nitem*Prox*Type*S	36	224.24		
Preso*Prox*Type	2	27.27	0.28	0.7619
Preso*Prox*Type*S	18	98.75		
Ncla*Nitm*Pres*Prox	12	283.57	1.41	0.1716
Ncla*Nitm*Pres*Prox*S	108	200.87		
Ncla*Nitm*Pres*Type	6	158.53	0.66	0.6857
Ncla*Nitm*Pres*Type*S	54	241.91		
Ncla*Nitm*Prox*Type	12	128.84	0.53	0.8896
Ncla*Nitm*Prox*Type*S	108	242.13		
Ncla*Pres*Prox*Type	6	305.66	0.36	0.8334
Ncla*Pres*Prox*Type*S	54	264.20		
Nitm*Pres*Prox*Type	4	104.48	0.36	0.8334
Nitm*Pres*Prox*Type*S	36	287.93		
Ncla*Nit*Pre*Pro*Typ	12	270.30	1.53	0.1232
Ncla*NIt*Pre*Pro*Typ*S	108	176.28		
Subjects	9	81414.51		
Total	1439			

TABLE F2

ANOVA Summary Table for Error Rate (Experiment 1)

Source of Variance	df	MS	F	p
Number Of Classes (Nclass)	3	0.897222	10.87	0.0001
Nclass*Subjects	27	0.082579		
Number Of Items (Nitem)	2	0.127083	3.27	0.0615
Nitem*S	18	0.038863		
Presentation Order (Preso)	1	0.003704	0.12	0.7415
Preso*S	9	0.031996		
Proximity (Prox)	2	0.258333	5.19	0.0166
Prox*S	18	0.049743		
Type	1	0.112037	1.15	0.3107
Type*S	9	0.097119		
Nclass*Nitem	6	0.109491	2.30	0.0478
Nclass*Nitem*S	54	0.047677		
Nclass*Preso	3	0.188889	5.04	0.0067
Nclass*Preso*S	27	0.037483		
Nclass*Prox	6	0.014815	0.32	0.9211
Nclass*Prox*S	54	0.045593		
Nclass*Type	3	0.129321	2.59	0.0735
Nclass*Type*S	27	0.049931		
Nitem*Preso	2	0.169676	3.16	0.0666
Nitem*Preso*S	18	0.053678		
Nitem*Prox	4	0.039583	0.72	0.5853
Nitem*Prox*S	36	0.055144		
Nitem*Type	2	0.323843	4.50	0.0261
Nitem*Type*S	18	0.072042		
Preso*Prox	2	0.373148	7.89	0.0035
Preso*Prox*S	18	0.047274		
Preso*Type	1	0.003704	0.05	0.8316
Preso*Type*S	9	0.077263		
Prox*Type	2	0.034259	0.79	0.4681
Prox*Type*S	18	0.047274		
Nclass*Nitem*Preso	6	0.011343	0.21	0.9723
Nclass*Nitem*Preso*S	54	0.054124		
Nclass*Nitem*Prox	12	0.035417	0.71	0.7421
Nclass*Nitem*Prox*S	108	0.050120		
Nclass*Nitem*Type	6	0.006867	0.15	0.9881
Nclass*Nitem*Type*S	54	0.045533		
Nclass*Preso*Prox	6	0.072222	2.14	0.0633
Nclass*Preso*Prox*S	54	0.033728		
Nclass*Preso*Type	3	0.134568	2.91	0.0527
Nclass*Preso*Type*S	27	0.046262		
Nclass*Prox*Type	6	0.067284	1.27	0.2865
Nclass*Prox*Type*S	54	0.052966		

TABLE F2 (continued)

Nitem*Preso*Prox	4	0.041204	0.51	0.7317
Nitem*Preso*Prox*S	36	0.081456		
Nitem*Preso*Type	2	0.001620	0.04	0.9622
Nitem*Preso*Type*S	18	0.042001		
Nitem*Prox*Type	4	0.039815	0.74	0.5712
Nitem*Prox*Type*S	36	0.053832		
Preso*Prox*Type	2	0.067593	0.93	0.4117
Preso*Prox*Type*S	18	0.072479		
Ncla*Nitm*Pres*Prox	12	0.031481	0.62	0.8234
Ncla*Nitm*Pres*Prox*S	108	0.050986		
Ncla*Nitm*Pres*Type	6	0.042670	1.08	0.3859
Ncla*Nitm*Pres*Type*S	54	0.039498		
Ncla*Nitm*Prox*Type	12	0.017747	0.40	0.9594
Ncla*Nitm*Prox*Type*S	108	0.043939		
Ncla*Pres*Prox*Type	6	0.012346	0.19	0.9786
Ncla*Pres*Prox*Type*S	54	0.065243		
Nitm*Pres*Prox*Type	4	0.032176	0.76	0.5552
Nitm*Pres*Prox*Type*S	36	0.042078		
Ncla*Nit*Pre*Pro*Typ	12	0.049614	0.91	0.5400
Ncla*NIt*Pre*Pro*Typ*S	108	0.051751		
Subjects	9	0.252674		
Total	1439			

TABLE F3

ANOVA Summary Table for Response Time (Experiment 1)

Source of Variance	df	MS	F	p
Number Of Classes (Nclass)	3	11328322.67	46.04	0.0001
Nclass*Subjects	27	246061.59		
Number Of Items (Nitem)	2	142306.00	0.80	0.4651
Nitem*S	18	178130.34		
Presentation Order (Preso)	1	241397.65	2.82	0.1272
Preso*S	9	85503.35		
Proximity (Prox)	2	283828.82	1.09	0.3560
Prox*S	18	259344.61		
Type	1	22984065.05	17.32	0.0024
Type*S	9	1327184.13		
Nclass*Nitem	6	142161.90	1.06	0.3987
Nclass*Nitem*S	54	134289.51		
Nclass*Preso	3	185330.17	2.18	0.1131
Nclass*Preso*S	27	84866.96		
Nclass*Prox	6	338559.67	2.69	0.0232
Nclass*Prox*S	54	125728.81		
Nclass*Type	3	805457.51	3.17	0.0404
Nclass*Type*S	27	254164.01		
Nitem*Preso	2	487567.90	1.44	0.2626
Nitem*Preso*S	18	338214.00		
Nitem*Prox	4	162171.75	1.38	0.2604
Nitem*Prox*S	36	117542.13		
Nitem*Type	2	304994.48	2.12	0.1496
Nitem*Type*S	18	144180.50		
Preso*Prox	2	307839.29	2.30	0.1286
Preso*Prox*S	18	133614.93		
Preso*Type	1	215717.07	2.70	0.1350
Preso*Type*S	9	80006.71		
Prox*Type	2	33318.06	0.70	0.5116
Prox*Type*S	18	47888.12		
Nclass*Nitem*Preso	6	285684.02	2.33	0.0452
Nclass*Nitem*Preso*S	54	122763.00		
Nclass*Nitem*Prox	12	172736.73	1.06	0.4010
Nclass*Nitem*Prox*S	108	162969.33		
Nclass*Nitem*Type	6	289423.72	2.38	0.0409
Nclass*Nitem*Type*S	54	121501.27		
Nclass*Preso*Prox	6	147414.02	0.53	0.7814
Nclass*Preso*Prox*S	54	277016.82		

TABLE F3 (Continued)

Nclass*Preso*Type	3	76183.55	0.67	0.5790
Nclass*Preso*Type*S	27	114038.09		
Nclass*Prox*Type	6	252670.58	1.64	0.1537
Nclass*Prox*Type*S	54	153951.32		
Nitem*Preso*Prox	4	331344.04	1.61	0.1918
Nitem*Preso*Prox*S	36	205234.30		
Nitem*Preso*Type	2	21903.58	0.10	0.9041
Nitem*Preso*Type*S	18	216149.75		
Nitem*Prox*Type	4	72033.49	0.27	0.8928
Nitem*Prox*Type*S	36	262829.58		
Preso*Prox*Type	2	213238.47	0.57	0.5768
Preso*Prox*Type*S	18	375831.13		
Ncla*Nitm*Pres*Prox	12	74381.01	0.35	0.9772
Ncla*Nitm*Pres*Prox*S	108	212538.30		
Ncla*Nitm*Pres*Type	6	116314.05	0.80	0.5729
Ncla*Nitm*Pres*Type*S	54	145068.74		
Ncla*Nitm*Prox*Type	12	87988.85	0.72	0.7322
Ncla*Nitm*Prox*Type*S	108	122746.63		
Ncla*Pres*Prox*Type	6	93246.78	0.81	0.5664
Ncla*Pres*Prox*Type*S	54	115044.20		
Nitm*Pres*Prox*Type	4	204938.39	1.91	0.1298
Nitm*Pres*Prox*Type*S	36	107243.93		
Ncla*Nit*Pre*Pro*Typ	12	305831.32	2.27	0.0130
Ncla*NIt*Pre*Pro*Typ*S	108	134533.50		
Subjects	9	8622103.14		
Total	1439			

TABLE F4

Newman-Keuls Test of Number Of Classes, SWAT Rating (Experiment 1)

Number Of Classes	SWAT	
2	13.03	A
3	23.35	B
4	31.50	C
5	41.03	D

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F5

Newman-Keuls Test of Number Of Items, SWAT Rating (Experiment 1)

Number Of Items	SWAT	
1	19.65	A
3	29.06	B
5	33.40	C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F6

Newman-Keuls Test of Number Of Classes X Number Of Items, SWAT Rating (Experiment 1)

Number Of Classes	Number Of Items	Number Of Symbols	SWAT	
2	1	2	5.98	A
3	1	3	14.34	B
2	3	6	14.62	B
2	5	10	20.36	C
4	1	4	23.79	D
3	3	9	25.89	E
3	5	15	29.79	F
4	3	12	33.02	G
5	1	5	34.48	G
4	5	20	37.68	H
5	3	15	42.85	I
5	5	25	45.78	J

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F7

Newman-Keuls Test of Proximity, SWAT Rating (Experiment 1)

Proximity	SWAT	
Tight	21.84	A
Medium	27.17	B
Random	33.14	C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F8

Newman-Keuls Test of Presentation Order X Proximity, SWAT Rating (Experiment 1)

Presentation Order	Proximity	SWAT			
CIS	Tight	21.16	A		
CIS	Medium	21.17	A		
CIS	Random	22.28	A		
Parallel	Tight	22.51	A		
Parallel	Medium	33.17		B	
Parallel	Random	44.00			C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F9

Newman-Keuls Test of Number Of Items X Presentation Order, SWAT Rating (Experiment 1)

Number Of Items	Presentation Order	SWAT	
1	Tight	19.13	A
1	Medium	20.17	A
3	Random	22.51	A
5	Tight	22.98	A
3	Medium	35.69	B
5	Random	43.82	C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F10

Newman-Keuls Test of Number Of Items X Proximity, SWAT Rating (Experiment 1)

Number Of Items	Proximity	SWAT						
1	Medium	18.44	A					
1	Tight	20.15	A	B				
1	Random	20.35	A	B				
5	Tight	22.44		B				
3	Tight	22.92		B				
3	Medium	28.56			C			
5	Medium	34.51				D	E	
3	Random	35.81					E	
5	Random	43.26						F

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F11

Simple-Effect F-tests for Presentation Order X Proximity at Each Level of Number Of Items, SWAT Rating (Experiment 1)

Source of Variance	df	MS	F	p
Preso x Prox for Nitem = 1	2	400.72	0.82	> 0.05
Preso x Prox for Nitem = 3	2	19188.45	39.48	< 0.001
Preso x Prox for Nitem = 5	2	32982.95	67.86	< 0.001
Preso x Prox x Nitem x Sub	36	486.07		

TABLE F12

Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 1, SWAT Rating (Experiment 1)

Presentation Order	Proximity	SWAT	
Parallel	Medium	18.28	A
CIS	Medium	18.61	A
CIS	Random	18.79	A
CIS	Tight	19.99	A
Parallel	Tight	20.32	A
Parallel	Random	21.92	A

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F13

Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 3, SWAT Rating (Experiment 1)

Presentation Order	Proximity	SWAT			
CIS	Medium	21.44	A		
CIS	Tight	22.90	A		
Parallel	Tight	22.94	A		
CIS	Random	23.18	A		
Parallel	Medium	35.69		B	
Parallel	Random	48.44			C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F14

Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 5, SWAT Rating (Experiment 1)

Presentation Order	Proximity	SWAT			
CIS	Tight	20.59	A		
CIS	Medium	23.47	A		
Parallel	Tight	24.28	A		
CIS	Random	24.87	A		
Parallel	Medium	45.54		B	
Parallel	Random	61.65			C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F15

Newman-Keuls Test of Number Of Items X Type, SWAT Rating (Experiment 1)

Number Of Items	Type	SWAT			
1	Positive	19.38	A		
1	Negative	19.92	A		
3	Positive	27.92		B	
5	Negative	30.27		B	C
3	Positive	31.40			C
5	Negative	35.40				D

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F16

Newman-Keuls Test of Number Of Classes, Error Rate (Experiment 1)

Number Of Classes	Error Rate			
2	0.0278	A		
3	0.0370		B	
4	0.0593		B	C
5	0.0926			C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F17

Newman-Keuls Test of Number Of Classes X Number Of Items, Error Rate (Experiment 1)

Number Of Classes	Number Of Items	Number Of Symbols	Error Rate	
2	1	2	0.0222	A
2	5	10	0.0278	A
3	3	9	0.0278	A
2	3	6	0.0333	A
3	1	3	0.0417	A
3	5	15	0.0417	A
4	3	12	0.0417	A
4	1	4	0.0444	A
5	1	5	0.0722	A
4	5	20	0.0917	A
5	5	25	0.0944	A
5	3	15	0.1111	B

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F18

Newman-Keuls Test of Proximity, Error Rate (Experiment 1)

Proximity	Error Rate	
Tight	0.0444	A
Medium	0.0486	A
Random	0.0694	B

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F19

Newman-Keuls Test of Presentation Order X Proximity, Error Rate (Experiment 1)

Presentation Order	Proximity	Error Rate		
Parallel	Tight	0.0347	A	
Parallel	Medium	0.0417	A	
CIS	Random	0.0500	A	
CIS	Tight	0.0542	A	
CIS	Medium	0.0556	A	
Parallel	Random	0.0889		B

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F20

Newman-Keuls Test of Number of Class X Presentation Order, Error Rate (Experiment 1)

Number Of Classes	Presentation Order	Error Rate		
2	Parallel	0.0148	A	
3	CIS	0.0352	A	
3	Parallel	0.0389	A	
2	CIS	0.0407	A	
4	CIS	0.0407	A	
4	Parallel	0.0778		B
5	Parallel	0.0889		B
5	CIS	0.0961		B

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F21

Newman-Keuls Test of Number Of Items X Type, Error Rate (Experiment 1)

Number Of Items	Type	Error Rate		
1	Negative	0.0361	A	
5	Positive	0.0431	A	
3	Positive	0.0500	A	
1	Positive	0.0542	A	
3	Negative	0.0569	A	
5	Negative	0.0847		B

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE F22

Newman-Keuls Test of Number Of Classes, Response Time (Experiment 1)

Number Of Classes	Response Time				
2	726.61	A			
3	804.57		B		
4	883.12			C	
5	964.85				D

Note: Means with the same letters are not significantly different, $p > 0.05$.

APPENDIX G

PARTICIPANTS' INSTRUCTIONS FOR EXPERIMENT 2

PARTICIPANTS' INSTRUCTIONS FOR EXPERIMENT 2

Symbol Population.

Sixty-four symbols have been divided among eight symbol classes. study these symbols so that you will be able to quickly identify each symbol and assign it to its correct class.

Your Task.

In each experimental session you will perform a series of 144 memory trials. Each trial will require completion of the following steps:

1.) A total of from 2 to 25 symbols will be displayed on the computer screen. You should store (in memory) the names of as many of these symbols as possible.

2.) Next a test symbol will be presented. The test symbol is called the *probe*. Your job is to indicate whether the probe was or was not one of the symbols previously shown. You will press the "yes" or the "no" key to indicate your response. Accuracy of response is much more important than speed of response.

3.) I will be changing the spatial arrangement of the displayed symbols and also the timing (whether symbols are presented all at once or sequentially). After each trial, you will rate how easy or difficult you thought it was to solve that problem given the display used. You should think about your rating carefully.

Practice.

Now, I will guide you through 10 practice trials to help you understand these instructions.

Remember.

Accuracy of response is much more important than speed of response.

APPENDIX H

TABLES FOR EXPERIMENT 2

TABLE H1

ANOVA Summary Table for SWAT Rating (Experiment 2)

Source of Variance	df	MS	F	p
Number Of Classes (Nclass)	3	116525.84	118.26	0.0001
Nclass*Subjects	27	985.31		
Number Of Items (Nitem)	2	836743.17	99.83	0.0001
Nitem*S	18	8381.80		
Presentation Order (Preso)	1	74366.34	20.86	0.0014
Preso*S	9	3564.18		
Proximity (Prox)	2	15217.00	29.41	0.0001
Prox*S	18	517.46		
Type	1	2730.35	5.75	0.0400
Type*S	9	474.53		
Nclass*Nitem	6	5140.93	9.96	0.0001
Nclass*Nitem*S	54	516.00		
Nclass*Preso	3	4901.28	9.34	0.0002
Nclass*Preso*S	27	524.93		
Nclass*Prox	6	422.06	1.77	0.1221
Nclass*Prox*S	54	237.99		
Nclass*Type	3	142.07	0.46	0.7128
Nclass*Type*S	27	309.10		
Nitem*Preso	2	6963.23	6.28	0.0085
Nitem*Preso*S	18	1109.36		
Nitem*Prox	4	3814.66	11.93	0.0001
Nitem*Prox*S	36	319.70		
Nitem*Type	2	77.92	0.33	0.7220
Nitem*Type*S	18	234.89		
Preso*Prox	2	9504.50	22.84	0.0001
Preso*Prox*S	18	416.17		
Preso*Type	1	81.90	0.35	0.5712
Preso*Type*S	9	237.09		
Prox*Type	2	295.90	1.27	0.3049
Prox*Type*S	18	233.04		
Nclass*Nitem*Preso	6	245.54	0.97	0.4521
Nclass*Nitem*Preso*S	54	253.26		
Nclass*Nitem*Prox	12	133.94	0.59	0.8457
Nclass*Nitem*Prox*S	108	226.78		
Nclass*Nitem*Type	6	129.81	1.05	0.4032
Nclass*Nitem*Type*S	54	123.49		
Nclass*Preso*Prox	6	232.54	0.80	0.5717
Nclass*Preso*Prox*S	54	289.48		

TABLE H1 (continued)

Nclass*Preso*Type	3	5.35	0.05	0.9865
Nclass*Preso*Type*S	27	115.45		
Nclass*Prox*Type	4	10.32	0.67	0.6704
Nclass*Prox*Type*S	36	170.70		
Nitem*Preso*Prox	4	1818.87	5.27	0.0019
Nitem*Preso*Prox*S	36	344.88		
Nitem*Preso*Type	2	85.98	0.35	0.7128
Nitem*Preso*Type*S	18	249.21		
Nitem*Prox*Type	4	10.32	0.06	0.9929
Nitem*Prox*Type*S	36	62.96		
Preso*Prox*Type	4	10.32	2.00	0.1647
Preso*Prox*Type*S	36	170.70		
Ncla*Nitm*Pres*Prox	12	248.45	1.32	0.2168
Ncla*Nitm*Pres*Prox*S	108	187.99		
Ncla*Nitm*Pres*Type	6	208.87	0.81	0.5639
Ncla*Nitm*Pres*Type*S	54	256.66		
Ncla*Nitm*Prox*Type	12	62.96	0.34	0.9793
Ncla*Nitm*Prox*Type*S	108	184.05		
Ncla*Pres*Prox*Type	6	414.41	1.57	0.1739
Ncla*Pres*Prox*Type*S	54	264.03		
Nitm*Pres*Prox*Type	4	551.40	2.45	0.0634
Nitm*Pres*Prox*Type*S	36	224.75		
Ncla*Nit*Pre*Pro*Typ	12	192.89	1.37	0.1927
Ncla*NIt*Pre*Pro*Typ*S	108	141.05		
Subjects	9	52828.88		
Total	1439			

TABLE H2

ANOVA Summary Table for Error Rate (Experiment 2)

Source of Variance	df	MS	F	p
Number Of Classes (Nclass)	3	1.464198	15.58	0.0001
Nclass*Subjects	27	0.093999		
Number Of Items (Nitem)	2	8.244676	76.09	0.0001
Nitem*S	18	0.108359		
Presentation Order (Preso)	1	0.045371	0.41	0.5380
Preso*S	9	0.110700		
Proximity (Prox)	2	0.067593	0.43	0.6599
Prox*S	18	0.127001		
Type	1	0.408333	0.44	0.5241
Type*S	9	0.929424		
Nclass*Nitem	6	0.199614	2.62	0.0268
Nclass*Nitem*S	54	0.076329		
Nclass*Preso	3	0.040432	0.35	0.7906
Nclass*Preso*S	27	0.116049		
Nclass*Prox	6	0.031790	0.36	0.9021
Nclass*Prox*S	54	0.088803		
Nclass*Type	3	0.205864	1.82	0.1673
Nclass*Type*S	27	0.113100		
Nitem*Preso	2	0.140509	2.81	0.0866
Nitem*Preso*S	18	0.049974		
Nitem*Prox	4	0.063773	0.82	0.5219
Nitem*Prox*S	36	0.077919		
Nitem*Type	2	0.063194	0.35	0.7126
Nitem*Type*S	18	0.183050		
Preso*Prox	2	0.017593	0.12	0.8888
Preso*Prox*S	18	0.145422		
Preso*Type	1	0.014815	0.14	0.7170
Preso*Type*S	9	0.105864		
Prox*Type	2	0.011111	0.19	0.8264
Prox*Type*S	18	0.057665		
Nclass*Nitem*Preso	6	0.078164	1.03	0.4180
Nclass*Nitem*Preso*S	54	0.076106		
Nclass*Nitem*Prox	12	0.069637	0.64	0.8002
Nclass*Nitem*Prox*S	108	0.108132		
Nclass*Nitem*Type	6	0.044059	0.43	0.8529
Nclass*Nitem*Type*S	54	0.101500		
Nclass*Preso*Prox	6	0.142284	1.72	0.1347
Nclass*Preso*Prox*S	54	0.082870		

TABLE H2 (continued)

Nclass*Preso*Type	3	0.083951	1.04	0.3924
Nclass*Preso*Type*S	27	0.081036		
Nclass*Prox*Type	6	0.056790	1.11	0.3694
Nclass*Prox*Type*S	54	0.051217		
Nitem*Preso*Prox	4	0.086690	0.84	0.5062
Nitem*Preso*Prox*S	36	0.102636		
Nitem*Preso*Type	2	0.112731	1.70	0.2113
Nitem*Preso*Type*S	18	0.066435		
Nitem*Prox*Type	4	0.033681	0.42	0.7904
Nitem*Prox*Type*S	36	0.079462		
Preso*Prox*Type	2	0.028704	0.26	0.7727
Preso*Prox*Type*S	18	0.109722		
Ncla*Nitm*Pres*Prox	12	0.084529	1.04	0.4144
Ncla*Nitm*Pres*Prox*S	108	0.080928		
Ncla*Nitm*Pres*Type	6	0.102238	1.26	0.2926
Ncla*Nitm*Pres*Type*S	54	0.081319		
Ncla*Nitm*Prox*Type	12	0.082137	1.01	0.4406
Ncla*Nitm*Prox*Type*S	108	0.080937		
Ncla*Pres*Prox*Type	6	0.123765	1.08	0.3879
Ncla*Pres*Prox*Type*S	54	0.114935		
Nitm*Pres*Prox*Type	4	0.063079	1.01	0.4139
Nitm*Pres*Prox*Type*S	36	0.062307		
Ncla*Nit*Pre*Pro*Typ	12	0.059066	0.65	0.7949
Ncla*NIt*Pre*Pro*Typ*S	108	0.090874		
Subjects	9	0.443640		
Total	1439			

TABLE H3

ANOVA Summary Table for Response Time (Experiment 2)

Source of Variance	df	MS	F	p
Number Of Classes (Nclass)	3	23862075.87	12.26	0.0001
Nclass*Subjects	27	1946966.63		
Number Of Items (Nitem)	2	178928785.38	25.22	0.0001
Nitem*S	18	7093848.22		
Presentation Order (Preso)	1	615807.89	0.51	0.4952
Preso*S	9	1219002.15		
Proximity (Prox)	2	424060.22	0.77	0.4775
Prox*S	18	550456.06		
Type	1	20578285.41	10.18	0.0110
Type*S	9	2021004.66		
Nclass*Nitem	6	340923.98	0.35	0.9073
Nclass*Nitem*S	54	976044.87		
Nclass*Preso	3	1034487.66	2.54	0.0773
Nclass*Preso*S	27	407032.42		
Nclass*Prox	6	261060.07	0.50	0.8070
Nclass*Prox*S	54	524136.34		
Nclass*Type	3	1072058.22	1.93	0.1487
Nclass*Type*S	27	555681.67		
Nitem*Preso	2	1190009.50	3.44	0.0544
Nitem*Preso*S	18	346213.72		
Nitem*Prox	4	459098.86	0.56	0.6964
Nitem*Prox*S	36	826685.06		
Nitem*Type	2	1429345.05	1.97	0.1686
Nitem*Type*S	18	725977.26		
Preso*Prox	2	1123005.80	2.02	0.1613
Preso*Prox*S	18	555245.23		
Preso*Type	1	27070.05	0.04	0.8474
Preso*Type*S	9	690460.07		
Prox*Type	2	156735.84	0.35	0.7106
Prox*Type*S	18	450106.06		
Nclass*Nitem*Preso	6	1919608.58	2.72	0.0220
Nclass*Nitem*Preso*S	54	704909.95		
Nclass*Nitem*Prox	12	486188.67	0.58	0.8555
Nclass*Nitem*Prox*S	108	840723.77		
Nclass*Nitem*Type	6	1015933.17	1.64	0.1543
Nclass*Nitem*Type*S	54	619850.94		
Nclass*Preso*Prox	6	643543.82	1.37	0.2447
Nclass*Preso*Prox*S	54	470813.59		

TABLE H3 (continued)

Nclass*Preso*Type	3	480424.94	0.64	0.5957
Nclass*Preso*Type*S	27	750380.25		
Nclass*Prox*Type	6	665652.55	0.88	0.5192
Nclass*Prox*Type*S	54	760427.94		
Nitem*Preso*Prox	4	928489.26	0.89	0.4819
Nitem*Preso*Prox*S	36	1047533.00		
Nitem*Preso*Type	2	1226053.87	1.11	0.3519
Nitem*Preso*Type*S	18	1107059.83		
Nitem*Prox*Type	4	705410.30	1.12	0.3629
Nitem*Prox*Type*S	36	630589.02		
Preso*Prox*Type	2	1943956.60	2.61	0.1009
Preso*Prox*Type*S	18	744141.75		
Ncla*Nitm*Pres*Prox	12	802416.65	1.79	0.0592
Ncla*Nitm*Pres*Prox*S	108	449032.02		
Ncla*Nitm*Pres*Type	6	357620.46	0.63	0.7060
Ncla*Nitm*Pres*Type*S	54	568218.07		
Ncla*Nitm*Prox*Type	12	1123512.01	1.60	0.1036
Ncla*Nitm*Prox*Type*S	108	704372.94		
Ncla*Pres*Prox*Type	6	947310.89	1.89	0.0993
Ncla*Pres*Prox*Type*S	54	501145.54		
Nitm*Pres*Prox*Type	4	213816.98	0.26	0.8993
Nitm*Pres*Prox*Type*S	36	810727.08		
Ncla*Nit*Pre*Pro*Typ	12	420241.63	0.49	0.9156
Ncla*NIt*Pre*Pro*Typ*S	108	854187.54		
Subjects	9	102332815.76		
Total	1439			

TABLE H4

Newman-Keuls Test of Number Of Classes, SWAT Rating (Experiment 2)

Number Of Classes	SWAT	
2	22.04	A
3	30.76	B
4	38.66	C
5	46.21	D

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H5

Newman-Keuls Test of Number Of Items, SWAT Rating (Experiment 2)

Number Of Items	SWAT	
1	8.26	A
3	39.25	B
5	55.74	C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H6

Newman-Keuls Test of Number Of Classes X Number Of Items, SWAT Rating (Experiment 2)

Number Of Classes	Number Of Items	Number Of Symbols	SWAT Rating	
2	1	2	1.93	A
3	1	3	4.08	A
4	1	4	9.56	B
5	1	5	17.48	C
2	3	6	22.44	D
3	3	9	35.95	E
2	5	10	41.75	F
4	3	12	44.93	F
3	5	15	52.25	G
5	3	15	53.67	G
4	5	20	61.48	H
5	5	25	67.49	I

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H7

Newman-Keuls Test of Proximity, SWAT Rating (Experiment 2)

Proximity	SWAT	
Tight	31.48	A
Medium	33.86	B
Random	37.91	C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H8

Newman-Keuls Test of Presentation Order X Proximity, SWAT Rating (Experiment 2)

Presentation Order	Proximity	SWAT				
CIS	Medium	29.61	A			
CIS	Tight	29.95	A			
CIS	Random	31.24	A	B		
Parallel	Tight	33.00		B		
Parallel	Medium	38.12			C	
Parallel	Random	44.57				D

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H9

Simple-Effect F-tests for Presentation Order at each Number Of Classes, SWAT Rating (Experiment 2)

Source of Variance	df	MS	F	p
Preso for Nclass = 2	1	425.1273	8.07	< 0.001
Preso for Nclass = 3	1	8990.8141	17.13	< 0.001
Preso for Nclass = 4	1	27165.2521	51.75	< 0.001
Preso for Nclass = 5	1	48668.9815	92.71	< 0.001
Preso x Nclass x Sub	27	524.9306		

TABLE H10

Simple-Effect F-tests for Presentation Order at each Number Of Items, SWAT Rating (Experiment 2)

Source of Variance	df	MS	F	p
Preso for Nitem = 1	1	3944.1960	3.56	> 0.5
Preso for Nitem = 3	1	35758.4067	32.23	< 0.001
Preso for Nitem = 5	1	48590.1938	43.80	< 0.001
Preso x Nitem x Sub	18	1109.3569		

TABLE H11

Simple-Effect F-tests for Proximity at each Number Of Items, SWAT Rating (Experiment 2)

Source of Variance	df	MS	F	p
Prox for Nitem = 1	2	56.6932	0.18	> 0.5
Prox for Nitem = 3	2	6084.4633	19.03	< 0.001
Prox for Nitem = 5	2	16705.1535	52.25	< 0.001
Prox x Nitem x Sub	36	319.7010		

TABLE H12

Simple-Effect F-tests for Number Of Items X Presentation Order X Proximity, SWAT Rating (Experiment 2)

Source of Variance	df	MS	F	p
Preso x Prox for Nitem = 1	2	87.5218	0.25	> 0.5
Preso x Prox for Nitem = 3	2	5292.3164	15.35	< 0.001
Preso x Prox for Nitem = 5	2	7762.4003	22.51	< 0.001
Nitem x Preso x Prox x Sub	36	344.8847		

TABLE H13

Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 1, SWAT Rating (Experiment 2)

Presentation Order	Proximity	SWAT	
CIS	Medium	6.52	A
CIS	Random	6.59	A
CIS	Tight	6.71	A
Parallel	Tight	9.04	A
Parallel	Medium	10.26	A
Parallel	Random	10.45	A

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H14

Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 3, SWAT Rating (Experiment 2)

Presentation Order	Proximity	SWAT	
CIS	Medium	33.63	A
CIS	Tight	34.41	A
CIS	Random	34.74	A
Parallel	Tight	38.05	A
Parallel	Medium	43.03	B
Parallel	Random	51.61	C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H15

Newman-Keuls Test of Presentation Order X Proximity at Number Of Items Equals 5, SWAT Rating (Experiment 2)

Presentation Order	Proximity	SWAT			
CIS	Medium	48.68	A		
CIS	Tight	48.72	A		
Parallel	Tight	51.93	A		
CIS	Random	52.40	A		
Parallel	Medium	61.07		B	
Parallel	Random	71.65			C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H16

Newman-Keuls Test Number Of Classes, Error Rate (Experiment 2)

Number Of Classes	Error Rate			
2	0.0685	A		
3	0.1000		B	
4	0.1111		B	
5	0.1574			C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H17

Newman-Keuls Test of Number Of Items, Error Rate (Experiment 2)

Number Of Items	Error Rate			
1	0.0287	A		
3	0.1208		B	
5	0.1785			C

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H18

Newman-Keuls Test of Number Of Classes X Number Of Items, Error Rate (Experiment 2)

Number Of Classes	Number Of Items	Number Of Symbols	Error Rate						
3	1	3	0.0111	A					
4	1	4	0.0194		B				
2	1	2	0.0250			C			
5	1	5	0.0583				D		
2	3	6	0.0583					E	
4	3	12	0.1194						F
2	5	10	0.1222						F
3	3	9	0.1278						F
3	5	15	0.1611						F	G
5	3	15	0.1778						F	G
4	5	20	0.1944							G	H
5	5	25	0.2361								H

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H19

Newman-Keuls Test of Number Of Classes, Response Time (Experiment 2)

Number Of Classes	Response Time (ms)	
2	1091.63	A
3	1212.86	A
4	1340.34	B
5	1432.04	B

Note: Means with the same letters are not significantly different, $p > 0.05$.

TABLE H20

Newman-Keuls Test of Number Of Items, Response Time (Experiment 2)

Number Of Items	Response Time (ms)		
1	872.48	A	
3	1388.82	A	
5	1546.36		B

Note: Means with the same letters are not significantly different, $p > 0.05$.

VITA

Michael Joel Kahn

May, 1990

Michael Kahn was born on May 12, 1957 in York, Pennsylvania. He attended Bucknell University and received a B.A. in English in 1979. In 1982 he earned an M.S. in psychology, human factors option, from North Carolina State University. His thesis assessed detection of fire-related cues by sleeping individuals.

Michael worked as a human factors engineer for FMC Corporation from 1982 to 1987. There he designed and evaluated numerous crew interfaces for ground combat vehicles.

He began dissertation work at Virginia Polytechnic Institute and State University in 1987. At Virginia Tech, he worked nine months as a research assistant for the Department of Computer Science. There he applied verbal protocol and impact analysis towards designing the user interface of a large application program (i.e., a user interface management system). Michael then worked 15 months with the Displays and Controls Laboratory (in the Department of Industrial Engineering and Operations Research) under the direction of Dr. Robert J. Beaton. He received his Ph.D. in industrial engineering and operations research, human factors engineering option, in May, 1990.